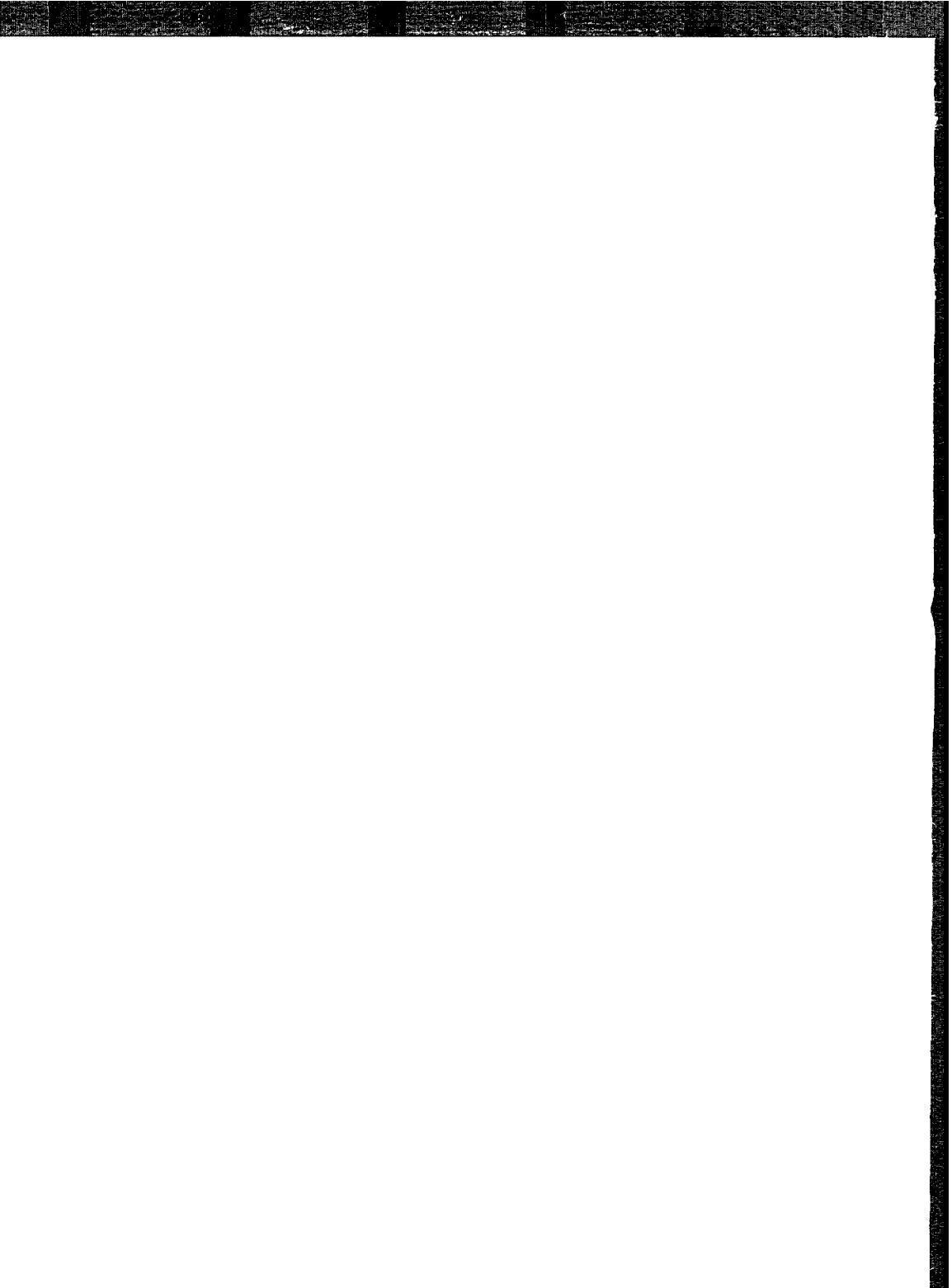




DISTRIBUTION STATEMENT A  
Approved for public release  
Distribution Unlimited

19950714 011



*This photograph of the Earth was taken from the Apollo 16 Spacecraft. Much of the Earth is heavily cloud covered. A portion of the United States from the Great Lakes to Southern California, including the Rocky Mountain area, is visible. The North American coastline from Southern Mexico to Alaska can be seen. (Courtesy of NASA.)*

~~USGCRP~~

Rept. no. USGCRP-95-02

# The U.S. Global Change Data and Information System Implementation Plan

Accession For	
NTIS	CRA&I <input checked="" type="checkbox"/>
DTIC	TAB <input type="checkbox"/>
Unannounced <input type="checkbox"/>	
Justification	
By <u>Judy Haller - DTIC-OS</u>	
Distribution 7-17-95	
Availability Codes	
Dist	Avail and / or Special
A-1	

A Report by the  
Committee on Environment and Natural Resources Research

1994

DTIC QUALITY INSPECTED 5

*The U.S. Global Change Data and Information System Implementation Plan* has been printed on recycled paper, May 1995.

For additional information on GCDIS or the USGCRP activities, or to obtain a copy of this document, contact:

Global Change Research Information Office

Gerald S. Barton  
Director, GCRC  
1825 K Street, N.W.  
Suite 805  
Washington, D.C. 20006  
USA  
(202) 775-6607  
(202) 775-6622 FAX  
[gbarton@gcrio.org](mailto:gbarton@gcrio.org) (e-mail)

GCRC User Services  
2250 Pierce Road  
University Center, MI 48710  
USA  
(517) 797-2730  
(517) 787-2622 FAX  
[help@gcrio.org](mailto:help@gcrio.org) (e-mail)

GCRC information services on the Internet:

GCRC Gopher service: [gcrio.org](http://gcrio.org)  
World Wide Web service: <http://www.gcrio.org>

THE WHITE HOUSE

WASHINGTON

February 14, 1995

I am pleased to release "The U.S. Global Change Data and Information System (GCDIS) Implementation Plan," a report of the Subcommittee on Global Change Research.

The importance of effective management of data and information on global change has been highlighted by the U.S. Global Change Research Program. Global change research activities generate and require massive amounts of highly diverse data and information. These data and information are needed to document change, to improve understanding of global change processes, and to carry out integrated assessments of impacts on human affairs. Because global change issues are so broad, including human and natural forcing factors, ecological change, biodiversity, human interactions, and comprehensive assessments, a data and information management challenge is to archive, preserve, and make data available in a useful form for researchers and other stakeholders. To help meet this requirement, a coordinated interagency Global Change Data and Information System (GCDIS) is being initiated.

GCDIS coordination is between all the applicable elements of the federal government as well as with GCDIS's national and international user communities. This user community spans the full range from K-12 educators to senior researchers and from governmental to non-governmental policy makers to public interest, commercial, and industrial organizations.

The conceptual design for GCDIS has been developed through a number of National Academy of Sciences studies, U.S. Global Change Research Program reports, national data policy principles adopted in 1991, and the 1992 GCDIS program plan that commits the agencies to its establishment. Building upon these accomplishments, this GCDIS Implementation Plan describes the structure the agencies will use to provide the multifaceted coordination required to meet the needs of the broad spectrum of GCDIS users.

I especially want to highlight the leadership of the Interagency Working Group on Data Management for Global Change (IWGDMGC) for developing and drafting this implementation plan. Additionally, this Plan required the active help and support of all the participating agencies and of the Committee on Geophysical and Environmental Data of the National Academy of Sciences.

Sincerely,



John H. Gibbons

Assistant to the President

for

Science and Technology

**OFFICE OF SCIENCE AND TECHNOLOGY POLICY  
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL**

**COMMITTEE ON ENVIRONMENT AND  
NATURAL RESOURCES RESEARCH (CENR)**

**FULL COMMITTEE**

*° Denotes Point of Contact*

**Co-Chairs**

D. James Baker	Department of Commerce, National Oceanic and Atmospheric Administration
H. Ronald Pulliam	Department of Interior
° Paul Dresler	Department of Interior
Thomas Lovejoy	Smithsonian Institution
Robert Watson	Office of Science and Technology Policy
° Rosina Bierbaum	Office of Science and Technology Policy

**Vice Chairs**

Robert Sussman	Environmental Protection Agency
Christine Ervin	Department of Energy

**NSTC Liaison**

Joanne Rodman	Office of Science and Technology Policy
---------------	---

**Executive Secretariat**

Susan Fruchter	Department of Commerce, National Oceanic and Atmospheric Administration
Donna Wieting	Department of Commerce, National Oceanic and Atmospheric Administration

**NSTC Executive Secretary**

Angela Philips Diaz	National Science and Technology Council
Laurel Kayse	National Science and Technology Council

**Office of Management and Budget**

T.J. Gauthier
Kathy Peroff
Jack Fellows

## **EXECUTIVE COMMITTEE**

*° Denotes Point of Contact*

D. James Baker	Department of Commerce, National Oceanic and Atmospheric Administration
H. Ronald Pulliam	Department of Interior
° Paul Dresler	Department of Interior
Tom Lovejoy	Smithsonian Institution
Robert Watson	Office of Science and Technology Policy
° Rosina Bierbaum	Office of Science and Technology Policy
Robert Sussman	Environmental Protection Agency
Christine Ervin	Department of Energy
Katie McGinty	Office of Environmental Programs
Joe Stiglitz	Council of Economic Advisors
° Sally Kane	Council of Economic Advisors
T.J. Glauthier	Office of Management and Budget

### **Agency Representatives Nominated To Full Committee**

Michael Stegman	Department of Housing and Urban Development
Jim Lyons	United States Department of Agriculture
° Allison Biggs	United States Department of Agriculture
Dr. Joseph Osterman	Department of Defense
° CAPT Bradley P. Smith	Department of Defense
Dr. Debra Knopman	Department of Interior
° Paul Dresler	Department of Interior
Dr. Anne Petersen	National Science Foundation
° Penny Firth	National Science Foundation
Elinor G. Constable	Department of State
° William Sullivan	Department of State
Dr. Charles Kennel	National Aeronautics and Space Administration
° Dr. Nancy Maynard	National Aeronautics and Space Administration
Dr. Kenneth Olden	Department of Health and Human Services
° Mary Gant	National Institute of Environmental Health Services
Bruce Burke	Domestic Policy Council
Robert H. Volland	Federal Emergency Management Agency
Dr. Ronald L. Ritschard	Tennessee Valley Authority
W. Bowman Cutter	National Economic Council
Susan Tierney	Department of Energy
° Jessie Harris	Department of Energy
Joseph Canny	Department of Transportation
Linda Zall	Central Intelligence Agency
Julian Skip Wright	Office of the Federal Coordinator for Meteorological Services

**COMMITTEE ON ENVIRONMENT AND NATURAL RESOURCES RESEARCH  
SUBCOMMITTEE ON GLOBAL CHANGE RESEARCH (SGCR)**

Chair

Robert W. Corell National Science Foundation

## Vice Chairs

## **Working Groups**

### Observations

## **Chairs and Vice Chairs**

Nancy Maynard (Chair), National Aeronautics and Space Administration

Thomas Nelson (Vice Chair), Department of Defense

Dixon Butler (Chair), National Aeronautics and Space Administration

Greg Withee (Vice Chair), Department of Commerce,  
National Oceanic and Atmospheric Administration

Aristides Patrinos (Chair), Department of Energy

Michael Carr (Vice Chair), Department of the Interior  
Michael Reeve (Vice Chair), National Science Foundation  
L. Michael Schlesinger

J. Michael Hall (Chair), Department of Commerce,  
National Oceanic and Atmospheric Administration  
Jeffrey E. Givens (Vice Chair), National Science Foundation

Jay Fein (Vice Chair), National Science Foundation  
Courtney Riordan (Chair), Environmental Protection Agency

Agency  
Mary Gant (Vice Chair), Department of Health and  
Human Services, National Institute of Environmental  
Health Sciences

Ted Maxwell (Vice Chair), Smithsonian Institution  
Abraham Haspel (Chair), Department of Energy  
Robin Cantor (Vice Chair), National Science Foundation

## Cross-Cutting Working Groups

## International Coordination/ Development Education and Community

Louis B. Brown (Chair), National Science Foundation  
Jonathan Pershing (Vice Chair), Department of State  
Gary R. Evans (Chair), U.S. Department of Agriculture  
Jane Stutsman (Vice Chair), National Science Foundation  
John Kermond (Vice Chair), Department of Commerce,  
National Oceanic and Atmospheric Administration

**Additional Members**

Thomas J. Baerwald	National Science Foundation
Truman Goins	Department of Housing and Urban Development
Kevin O'Connell	Intelligence Community
Thomas Rhodes	U.S. Agency for International Development
Heather Ross	National Economic Council
Fred Saalfeld	Department of Defense
William L. Sullivan, Jr.	Department of State
Frank Thornton	Tennessee Valley Authority

**Alternates**

Pete Comanor	Department of the Interior
Roger Dahlman	Department of Energy
James Fouts	Department of Health and Human Services
Richard Greenfield	National Science Foundation
Jeffrey M. Maclure	Department of State
Norris Nielsen	Tennessee Valley Authority
Rosemarie Russo	Environmental Protection Agency
Dennis Tirpak	Environmental Protection Agency
Carol Whitman	U.S. Department of Agriculture

**Office of the U.S. Global Change Research Program**

Michael MacCracken, Director  
Sylvia Edgerton, Deputy Director  
Sean Downing  
Wanda Ferrell  
Rick Kelly  
Richard Moss  
Warren Spaeth

**Global Change Research Information Office**

Gerald Barton, Director

**INTERAGENCY WORKING GROUP ON  
DATA MANAGEMENT FOR GLOBAL CHANGE (IWGDMGC)**

**Principals**

James Biesecker	Department of the Interior, United States Geological Survey
Robert Corell	National Science Foundation
Charles Kennel	National Aeronautics and Space Administration
Ari Patrinos	Department of Energy
Courtney Riordan	Environmental Protection Agency
Norton Strommen	U.S. Department of Agriculture
Edward Whitman	Department of Defense, United States Navy
Greg Withee	Department of Commerce, National Oceanic and Atmospheric Administration, Chairman

**Liaisons**

Francis Bretherton	University of Wisconsin
Richard Hallgren	American Meteorological Society

**Contacts**

Pam Andre	Department of Agriculture
Gerald Barton	Department of Commerce, National Oceanic and Atmospheric Administration
Dorothy Bergamaschi	Department of State
Arthur Booth	Department of Commerce, National Oceanic and Atmospheric Administration
Dixon Butler	National Aeronautics and Space Administration
Claude Christensen	Department of the Interior
Eliot Christian	Department of the Interior, United States Geological Survey
Louis DeMouy	Department of Energy
Larry Enomoto*	Department of Commerce, National Oceanic and Atmospheric Administration
Gary Evans	U.S. Department of Agriculture

Mike Farrell*	Department of Energy, Oak Ridge National Laboratory
Jay Fein	National Science Foundation
Wanda Ferrell	Department of Energy
Greg Hunolt	National Aeronautics and Space Administration
Paul Kanciruk	Department of Energy, Oak Ridge National Laboratory
Barbara Levinson	Environmental Protection Agency
Thomas Mace	Environmental Protection Agency
Brand Niemann	Environmental Protection Agency
Larry Pettinger	Department of the Interior, United States Geological Survey
Jenny Plummer	Department of Defense, United States Navy
Robbie Rand*	U.S. Department of Agriculture
Dean Smehil	Department of Defense, United States Navy
Roger Taylor*	National Science Foundation
Jim Thieman*	National Aeronautics and Space Administration
Richard Thompson	U.S. Department of Agriculture
Carol Watts*	Department of Commerce, National Oceanic and Atmospheric Administration
Roxanne Williams	U.S. Department of Agriculture
Helen Wood	Department of Commerce, National Oceanic and Atmospheric Administration

**Executive Secretary**

Les Meredith	National Aeronautics and Space Administration
--------------	---

\* Subgroup Chair

# Table of Contents

Executive Summary .....	ix
1. Introduction.....	1
The Global Change Data and Information System.....	1
Scope .....	2
Implementation Principles.....	4
Measures of Success.....	6
2. Implementation Coordination .....	7
Program Management .....	7
GCDIS Coordination .....	11
External Coordination .....	13
Policy Coordination.....	13
User Coordination .....	14
3. Users and the GCDIS.....	16
Diversity of the Users.....	17
The Problem Domain of the Users .....	19
The Uses of the Data and Information by the User .....	20
Organizational and Group Affiliation of Users .....	22
Libraries and Information Centers as Intermediaries and Facilitators.....	23
Implications of the User Attributes on the GCDIS.....	24
User Support.....	25
4. Content System Implementation .....	27
Conceptual View of GCDIS Content .....	27
Scope of Content Subgroup.....	28
Coordination of Content Implementation.....	29
Data and Information Identification, Acquisition, and Priorities .....	32
Assembly .....	35
Production of Derived, Value-Added Data Products and Services .....	40
Supporting Functions .....	41
National Policy Issue Data and Information .....	44
Emerging Trends .....	45
5. Access System Implementation by Functional Areas.....	47
Access System Overview .....	47
Coordinating Access Implementation .....	51
Access Services .....	52
User Support and Outreach .....	56
Networks .....	57
Directory.....	60

Guide .....	64
Inventory .....	66
Browse.....	68
Order Placement.....	69
Archive Function.....	70
Distribution Functions.....	72
Billing and Related Administrative Functions .....	73
Information Functions and Services .....	74
State and Local Government Access.....	76
International Access .....	77
<b>6. Agency Implementation.....</b>	<b>79</b>
Programmatic Guidance and Milestones.....	79
Overall Schedule Goals.....	80
Agency Implementation .....	83
The Evolution of the GCDIS.....	88
<b>7. International Links and Coordination .....</b>	<b>90</b>
Existing International Data and Information Systems and Services .....	91
International Mechanisms for Planning, Coordinating, and Implementing .....	94
<b>Appendices .....</b>	<b>101</b>
<b>A. Data Capture .....</b>	<b>102</b>
Data and Information Sources.....	102
The Special Issue.....	103
Data Capture Activities .....	106
Fiscal Requirements .....	109
<b>B. Agency Implementation Overviews.....</b>	<b>110</b>
The Department of Commerce .....	110
Department of Defense.....	116
Department of Energy .....	119
Department of the Interior .....	123
Environmental Protection Agency .....	130
National Aeronautics and Space Administration .....	137
National Science Foundation .....	147
U.S. Department of Agriculture .....	151
<b>C. List of Acronyms and Abbreviations .....</b>	<b>155</b>

## Executive Summary

The United States Global Change Research Program (USGCRP) was established to observe, understand, and predict global change, and to make its results available for use in policy matters. Fundamental to the understanding and prediction of global change and its associated human impacts is the timely availability, in useful forms, of the data and information required to conduct the research, develop the models, and make the assessments required. Much of these data and information originates in many separate programs of agencies of the Federal Government.

Activities of the USGCRP are coordinated by the Committee on Environment and Natural Resources Research (CENR). Because data and information are of fundamental importance to the USGCRP, the Committee on Earth and Environmental Sciences, which has been superseded by the CENR, published the *U.S. Global Change Data and Information Management Program Plan* in 1992 in which the participating Federal agencies committed to work with each other, with academia, and with the international community to make it as easy as possible for researchers and others to access and use global change data and information. The program plan was based on guidance received during a number of years from the user community and especially from the Committee on Geophysical and Environmental Data of the National Academy of Sciences. This document builds on that program plan to define the implementation of a Global Change Data and Information System (GCDIS).

The GCDIS is the set of individual agency data and information systems supplemented by a minimal amount of crosscutting new infrastructure, and made interoperable by use of standards, common approaches, technology sharing, and data policy coordination. The GCDIS user community extends from global change researchers to other researchers, policymakers, educators, private industry, and private citizens. Through the GCDIS, this user community will be able to learn what data and information are available, have the key holdings available in useful forms with ready access, and be assured of their quality and continued availability. This *U.S. Global Change Data and Information System Implementation Plan* documents the specific interagency agreements and arrangements established to implement the GCDIS.

The individual agency components of the GCDIS are funded through their respective budgets. There will be neither a lead agency nor a separately funded budget for the GCDIS. Interagency GCDIS implementation coordination will be through the

---

Interagency Working Group on Data Management for Global Change (IWGDMGC), serving as the panel for the USGCRP Working Group on Data and Information Management. Where commonality or interoperability is essential for the GCDIS, the agencies will make decisions in concert.

Under this implementation plan, all data and information functional activities are the responsibility of individual agencies. The global change data and information management program adds the following multiagency capabilities:

- Coordinated response to the needs of the other management elements of the USGCRP and of its overarching themes;
- Full involvement of users in GCDIS development, operation, and management;
- Provision to users of a view of all priority data and information regardless of source (agency, university, state, or international), type (digital, analog, or text), or whether or not the data or information were obtained through the focused programs of the USGCRP;
- Provision to users of the ability to identify, search, view, order, and (where possible) take electronic delivery of all data managed under this plan through access to any single agency system;
- Coordination of archiving and data product production activities, including production of data and information products that combine multidisciplinary data sets, to provide both continuity and the highest possible quality; and
- Coordination of budget and system planning under the CENR.

Essential to these capabilities are the adoption and implementation of data and information management standards, adherence to common data and information policies governing user access, and the development and implementation of additional data and information system functionality necessary to provide interoperability of the multiple agency systems. The following principles guide the implementation.

### **Implementation Drivers**

- There will be consistency with USGCRP research and policy priorities.
- The implementation will be driven by research, policy, and other user requirements.
- User satisfaction and extent of use will be employed to assess the success of the system and to guide its evolution.
- Individual agency-managed implementation efforts within a framework for cooperation will be provided by the implementation plan.

---

- There will be an open and extensible system architecture that plans for evolution and provides pointers to all priority data and information that are available with user support.
- Multilevel, multipath access that recognizes the needs of individual user communities will be available to users.
- Easy access to data and information resources, such as those maintained by States, other countries, and national and international organizations, will be facilitated by the adoption of locator technology freely available on the international Internet.

### **Funding Assumptions**

- Resource requirements for the focused program activities of the individual agencies for data and information management have been included in USGCRP planning.
- A USGCRP program will be implemented to provide access through the GCDIS to large quantities of priority data and information obtained by programs outside the global change program.
- The National Aeronautics and Space Administration will seek funding to provide the information services to enable users to locate data and information across the full set of agency systems.
- The network interconnection of the agency systems and their connection to the user community will be provided by the High Performance Computing and Communications Program and the National Research and Education Network.
- The pace and level of the GCDIS implementation will be limited by the availability of resources in each of the participating agencies.

### **Implementation Approach**

- The GCDIS implementation will ensure that the data and information priorities for the necessary aspects of the USGCRP are established.
- The GCDIS will be built upon many other agency programs that are vital to the USGCRP as a source of data and information, as well as of systems infrastructure without degrading the services those programs are now providing.
- GCDIS implementation will use and build upon the agency infrastructure and technology developed to meet agency-specific requirements.

- Federal Information Processing Standards and national and international standards will be used wherever possible to promote interoperability, extensibility, and capability for evolution.
- Agencies will implement procedures to ensure that the priority data and information resulting from their funding of individual global change researchers are made available to the GCDIS in forms useful to the user community.
- Libraries and information centers, with their data and information resources, services, and standards, will be an integral part of the GCDIS.
- Global change data and information will be accessed electronically through the Internet and its successor network, and will be distributed by a variety of methods.

Each of the participating agencies will play a role appropriate to its agency mission, consistent with the funds available to it. Brief descriptions of each agency's role follow. Each agency will implement the GCDIS at its own pace.

**The Department of Agriculture (USDA)** will identify data and information among its holdings that are of value to the global research program and make these metadata available through the standard locator approach defined by the IWGDMGC.

**The Department of Commerce (DOC)** will provide user access to global National Oceanic and Atmospheric Administration (NOAA) and DOD environmental data and information of value to the USGCRP through the interoperable NOAA Data and Information System; it also will provide access to the census data and information.

**The Department of Defense (DOD)** will make unclassified and unrestricted data and information and limited metadata available to appropriate U.S. national archives; agencies receiving DOD data products will assume the responsibility for GCDIS functions (i.e., archive, directory, distribution, network, and user-support functions).

**The Department of Energy (DOE)** will provide access to data and information from its programs through a set of individual data and information systems focused on specific data bases to which interoperable functionality will be added.

**The Department of the Interior (DOI)** will provide access to data and information from its programs and other sources for which the DOI is custodian through standard locator technology among data and information systems of DOI bureaus. GCDIS interoperability beyond the locator level

initially will be provided at the U.S. Geological Survey's Earth Resources Observation Systems Data Center and may be added to other systems.

**The Department of State** will maintain and coordinate links with intergovernmental organizations and United Nations technical agencies having programs relevant to the GCDIS. A high-priority objective is to promote wide adoption of a policy of full and open sharing of global change data in multilateral forums, as well as in bilateral agreements.

**The Environmental Protection Agency (EPA)** will initially provide its data and information produced under the USGCRP through other participating agencies and then will develop a unified electronic data and information system to make other relevant data and information from agency activities available.

**The National Aeronautics and Space Administration (NASA)** will make all the data and information from its Mission to Planet Earth available on an interoperable basis through the Earth Observing System Data and Information System and will seek funding to develop, as needed, the interoperable functionality for implementation by all participating agencies. NASA will provide also for the data and information archiving and management, where needed, of socioeconomic data and information through its Socioeconomic Data and Applications Center.

**The National Science Foundation (NSF)**, from the National Center for Atmospheric Research, the Long-Term Ecological Research (LTER) program, and the Center for Ecological Analysis and Synthesis in Environmental Biology, will provide access to archived data and information through an existing data system to which interoperability will be added. An electronic data system for data and information from the LTER sites will be developed. Researchers funded by the NSF will be required to make available all global-change-relevant data and information developed in the performance of publishable research through one of the participating data systems.

GCDIS participation will be broadened to include other agencies with data and information needed for the USGCRP. To encourage such participation, the GCDIS is planned as a system that an agency can join with only a modest commitment. Later, the agency can work toward a fuller implementation of the information service objectives of the GCDIS. While the desirable level of enhanced participation will involve provision by the agency of the full suite of metadata appropriate for their data and information, the minimal requirement for agency participation will be

- Availability of data or information relevant to global change,
- Directory-level metadata describing these data or information,
- A credibly permanent commitment to maintaining these data and information and making them available,
- Interoperability with the rest of the GCDIS sufficient to make its directory-level information accessible from any other GCDIS entity, and
- User support for the data or information by somebody.

Because of the evolving nature of the USGCRP, of which the IWGDMGC is an integral part, this implementation plan will be updated as necessary to reflect program changes and accomplishments. With these updates, the implementation plan will provide the basis for individual agency implementation plans, some of which have already been initiated. By the activities included in this implementation plan, the resources of the agencies, the power of new technologies, and the capabilities of users and the international community will be brought to bear on providing the global change data and information needed by users to understand global processes critical to human well-being on Earth.

## 1. Introduction

The United States Global Change Research Program (USGCRP) was established to observe, understand, and predict global change, and to make its results available for use in policy matters. To meet this goal, the USGCRP researchers must investigate issues that extend across traditional disciplines. The data and information needed include both discipline-specific and assimilated data sets that are combinations of data and information from disciplines that range from the geophysical and biological sciences to economics and social sciences. These data and information will result from physical observations, sociological surveys, or model results, and will transcend agency and national boundaries. Thus, the data and information management component of the USGCRP is challenged to build upon existing data center structures to create a more comprehensive and versatile resource.

In 1992, the Committee on Earth and Environmental Sciences (CEES) published the *U.S. Global Change Data and Information Management Program Plan* to address these challenges. In that plan, agencies participating in the USGCRP committed program resources (consistent with their roles and missions) to the goal of an interagency global change data and information management program, consistent across agencies, that involves and supports university, international, and other user communities. This present document uses that plan to define the implementation of a Global Change Data and Information System (GCDIS). It defines the interagency mechanisms that will be established to coordinate the efforts of individual agencies involved in implementing the GCDIS, overall timetables, and agency-by-agency overviews with schedules of the implementation efforts. This GCDIS implementation plan will evolve through regular updates to reflect changing USGCRP research priorities, agency plans, technology advancements, program accomplishments, and the changing needs of global change data and information users.

## The Global Change Data and Information System

The goal of the U.S. Global Change Data and Information Management Program is to make it as easy as possible for researchers and others to have ready access to and use of global change data and information. To achieve this goal, the USGCRP agencies are organizing the GCDIS, which builds upon the mission resources and responsibilities of each agency and links the data and information services of the agencies to each other and to the users.

The use of the word *system* in GCDIS embraces the full range of people, infrastructure (e.g., hardware, software, networks, telephones, mail), and procedures for identifying, assembling, documenting, archiving, and disseminating that as a whole can provide users with data and information services.

The GCDIS is the set of individual agency data and information systems supplemented by a minimal amount of crosscutting new infrastructure, and made interoperable by use of standards, common approaches, technology sharing, and data policy coordination. Thus, portions of each agency's own data and information system will be that agency's component of the GCDIS, and the GCDIS will emerge through this coordinated interagency participation. Some agencies may arrange to have other agencies assume responsibility for selected aspects of their GCDIS-related functions. This is how Department of Defense (DOD) data will be made available to the GCDIS.

## **Scope**

The scope of the GCDIS implementation effort is described in this plan in terms of the user community it serves, its data and information content, and the access services it provides. Each of these will be discussed in more detail later in this plan.

## **User Community**

The primary users of the GCDIS are the global change researchers in agencies, academia, and the international community who conduct process research, integrative modeling and prediction investigations, and science assessment studies; researchers, policymakers, and educators who assess the state of global change and global change research to provide information for policy decisions; and the public.

The level of services, data, and information each class of user requires will vary significantly. For example, process researchers will more likely need large samples of raw data that may include combinations of data streams from several experiments or observational systems. Assessment studies will include the synthesis of the output of models that, for example, may characterize the world economy, atmospheric chemistry, ocean circulation, ecological systems, and social preferences and dynamics. Policymakers more likely will seek summarized and evaluated information, much of which will be in the form of text, graphs, and tables.

The GCDIS user community also extends to those agencies, programs, projects, individuals, and others that are sources of the data and information in the GCDIS and for which the GCDIS will provide archiving and distribution. The GCDIS services

required by each of the different types of users will differ, but have many aspects in common.

### **Data and Information Content**

The scope and priorities for GCDIS data and information content are driven by the data and information required to carry out current and planned USGCRP activities. The scope of data and information for the GCDIS comprises the Earth science data holdings (including in situ and satellite) of the agencies, selected analyzed and assimilated data sets, selected output from global change models, published documents, and socioeconomic data and information (including the areas of population, economic systems, and political systems and institutions) necessary for the study of human dimensions of global change. A large fraction of these data and information will come from contributing sources rather than from the focused global change research program.

The USGCRP encompasses the geophysical, geochemical, biological, social, and economic sciences and the research involved with integrated assessments of environmental policy issues. Of these areas of activity, however, only a subset of the geosciences has a tradition of institutional support for data and information services provided through data centers. The other disciplines, having no such traditions, presently lack the funding or other incentives to build this structural capability. This difference is reflected by the near-term implementation emphasis in this plan on geophysical data and geophysical data centers.

For the longer term, agencies face a significant task simply to identify the vast array of data and information they currently hold that would contribute toward a better understanding of the environment. Following this identification, based on the highest priority areas of interest, the task is to design and subsequently implement data and information services adequate to support the full breadth of the USGCRP. To start this process, the agencies plan to implement several pilot projects that are intended to broaden the scope of the GCDIS beyond its present emphasis on geophysical data and information.

### **Data and Information Access Services**

The scope of the data and information access services to be provided by the cooperative action of agency elements of the GCDIS includes user service procedures; generic and specialized user interfaces; a hierarchical structure with multiple points of access for identifying relevant data and information sets including directory, guide, and inventory information, library and information center resources,

and browse products; data dictionary; search and order functions; data formats for users; and standards for distribution media and user documentation.

## **International Context**

The GCDIS is designed to permit extension to international data and information holders, providers, and users. An important means to access international data and information will be provided initially through existing agency participation in the World Data Center system. This system, conceived and fostered by the International Council of Scientific Unions (ICSU), is responsible for archiving and exchanging important environmental data and information among almost all nations. A number of agency data centers associated with the implementation of the GCDIS currently has responsibility under this system for a broad range of important global change data and information. All GCDIS archive sites will be linked to agency data centers with World Data Center status.

## **Implementation Principles**

The data and information management program will be implemented in the context of the following set of high-level principles.

### **Implementation Drivers**

- There will be consistency with USGCRP research and policy priorities.
- The implementation will be driven by research, policy, and other user requirements.
- User satisfaction and extent of use will be employed to assess the success of the system and to guide its evolution.
- Individual agency-managed implementation efforts within a framework for cooperation will be provided by the implementation plan.
- There will be an open and extensible system architecture that plans for evolution and provides pointers to all priority data and information that are available with user support.
- Multilevel, multipath access that recognizes the needs of individual user communities will be available to users.
- Easy access to data and information resources, such as those maintained by States, other countries, and national and international organizations, will be facilitated by the adoption of locator technology freely available on the international Internet.

**Funding Assumptions**

- Resource requirements for the focused program activities of the individual agencies for data and information management have been included in USGCRP planning.
- A USGCRP program will be implemented to provide access through the GCDIS to large quantities of priority data and information obtained by programs outside the global change program.
- The National Aeronautics and Space Administration (NASA) will seek funding to provide the information services to enable users to locate data and information across the full set of agency systems.
- The network interconnection of the agency systems and their connection to the user community will be provided by the High Performance Computing and Communications Program (HPCC) and the National Research and Education Network (NREN).
- The pace and level of the GCDIS implementation will be limited by the availability of resources in each of the participating agencies.

**Implementation Approach**

- The GCDIS implementation will ensure that the data and information priorities for the necessary aspects of the USGCRP are established.
- The GCDIS will be built upon many other agency programs that are vital to the USGCRP as a source of data and information, as well as of systems infrastructure without degrading the services those programs are now providing.
- GCDIS implementation will use and build upon the agency infrastructure and technology developed to meet agency-specific requirements.
- Federal Information Processing Standards (FIPS) and national and international standards will be used wherever possible to promote interoperability, extensibility, and capability for evolution.
- Agencies will implement procedures to ensure that the priority data and information resulting from their funding of individual global change researchers are made available to the GCDIS in forms useful to the user community.
- Libraries and information centers, with their data and information resources, services, and standards, will be an integral part of the GCDIS.

- Global change data and information will be accessed electronically through the Internet and its successor network, and will be distributed by a variety of methods.

## Measures of Success

In order to help evaluate how well the GCDIS is meeting the data and information management needs of the USGCRP community, a set of criteria for measuring the success of the GCDIS has been established. Such means of evaluation are critical both for identifying and resolving problem areas and for helping guide the evolution of the GSDIS. The criteria themselves will evolve based on experience with their use and changes of the constraints and environment within which the GCDIS must function. The criteria are as follows:

• Satisfaction of users by each user category	To help assess users' perception of the GCDIS
• Rate of increase in the number of repeat users in each user category	To assess the rate of growth of the number of users who find the GCDIS important enough for their work that they continue to use it
• Fractional use of the GCDIS compared with other data and information sources by each user category	To help assess GCDIS coverage and the importance of the GCDIS to its user community
• Fraction of data and information that is relatable and interoperable	To help assess the contribution the GCDIS makes to the easy interdisciplinary use of its data and information
• Satisfaction of the other program elements of the USGCRP	To assess how well the GCDIS meets the programmatic needs of the USGCRP.

The user categories for these evaluations will evolve, but initially they will be research (physical, biological, human dimensions), assessment and policy, operations, education, and international.

## **2. Implementation Coordination**

Implementing the GCDIS elements will be a cooperative effort by the agencies involved, with individual agencies being responsible for managing their own implementation efforts within a framework of coordination by the agencies. The individual agency components of the program are funded through their respective budgets. There will be no lead agency or separately funded budget for program coordination. Program leadership for GCDIS implementation will be provided at two levels. At the highest level, program advocacy, evaluation and problem resolution will continue to be the domain of the Interagency Working Group on Data Management for Global Change (IWGDMGC), serving as the USGCRP Data and Information Working Group (see Figure 1). On a day-to-day basis, program monitoring and coordination with the research components of the USGCRP will be provided by the associate scientist for data located in the USGCRP program office.

Implementation of the GCDIS will be embedded in each agency's implementation plan or upgrade of its own data and information system. This GCDIS implementation plan will provide an overall view of agency efforts, showing how they fit within the overall program implementation. Individual agency implementation plans will contain specific details for the agency elements involved in GCDIS, including linkage with international programs and institutions. GCDIS implementation must be consistent with USGCRP research priorities as reflected in agency plans and with the required services that flow from those plans.

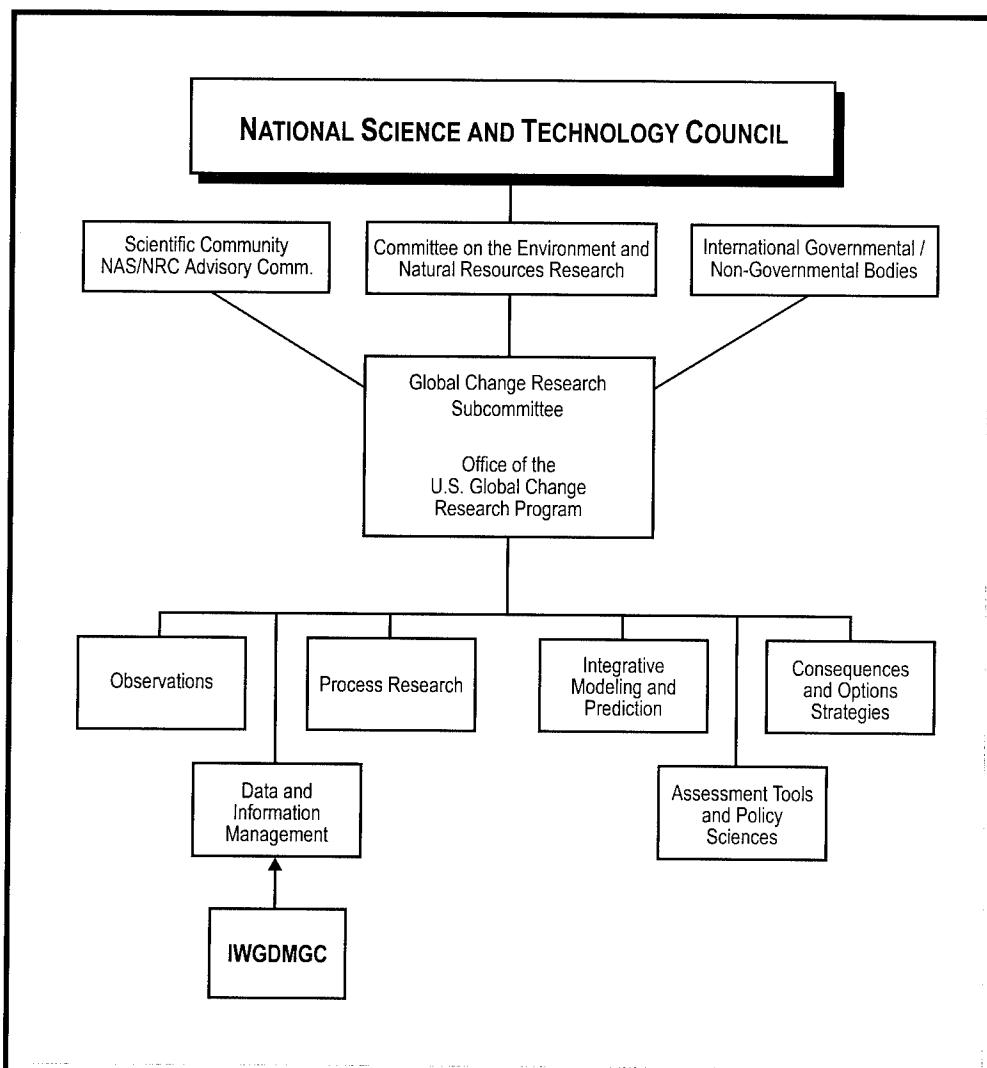
Closely synchronized implementation across the agencies is neither required nor practical. Agency missions vary, and not all agencies will be active in all of the GCDIS functional areas at the same time. Each agency may retain as much autonomy as is necessary. Where commonality or interoperability is essential for the GCDIS, decisions affecting the content or access system functionality will be taken by the IWGDMGC agencies in concert. The GCDIS will be open and extensible. This chapter describes the management and coordination of the program.

### **Program Management**

The Global Change Research Act of 1990 prescribed that the activities of the USGCRP were to be coordinated by the CEES. The CEES has been superseded by the Committee on Environment and Natural Resources Research (CENR). The data and information management component of the USGCRP reports to the CENR

Global Change Research Subcommittee through the Working Group on Observations and Data Management (see Figure 1). The global change data and information management program is coordinated at the working level by a standing group of agency representatives, one from each participating agency, a chairperson, and outside experts as needed. This group, the IWGDMGC, is responsible for program implementation and represents the agencies in budget coordination and planning for the data and information component of the USGCRP.

**Figure 1. Organization of the USGCRP**



Policies relevant to global change data and information will be reflected in individual agency implementation plans and collectively in the GCDIS implementation plan. In particular, the principles endorsed in the *Data Management for Global Change Research Policy Statements* (Table 1) provide a high-level framework for GCDIS implementation. Agencies participating in the USGCRP will develop corresponding internal policies and procedures that conform to the Data Policy Statements. The interagency coordination for GCDIS implementation will ensure that the collective efforts of the agencies result in the fulfillment of national policy requirements. For

**Table 1. Data Management for Global Change Research Policy Statements — July 1991**

- The U.S. Global Change Research Program requires an early and continuing commitment to the establishment, maintenance, validation, description, accessibility, and distribution of high-quality, long-term data sets.
- Full and open sharing of the full suite of global data sets for all global change researchers is a fundamental objective.
- Preservation of all data needed for long-term global change research is required. For each and every global change data parameter, there should be at least one explicitly designated archive. Procedures and criteria for setting priorities for data acquisition, retention, and purging should be developed by participating agencies, both nationally and internationally. A clearinghouse process should be established to prevent the purging and loss of important data sets.
- Data archives must include easily accessible information about the data holdings, including quality assessments, supporting ancillary information, and guidance and aids for locating and obtaining the data.
- National and international standards should be used to the greatest extent possible for media and for processing and communication of global data sets.
- Data should be provided at the lowest possible cost to global change researchers in the interest of full and open access to data. This cost should, as a first principle, be no more than the marginal cost of filling a specific user request. Agencies should act to streamline administrative arrangements for exchanging data among researchers.
- For those programs in which selected principal investigators have initial periods of exclusive data use, data should be made openly available as soon as they become widely useful. In each case, the funding agency should explicitly define the duration of any exclusive use period.

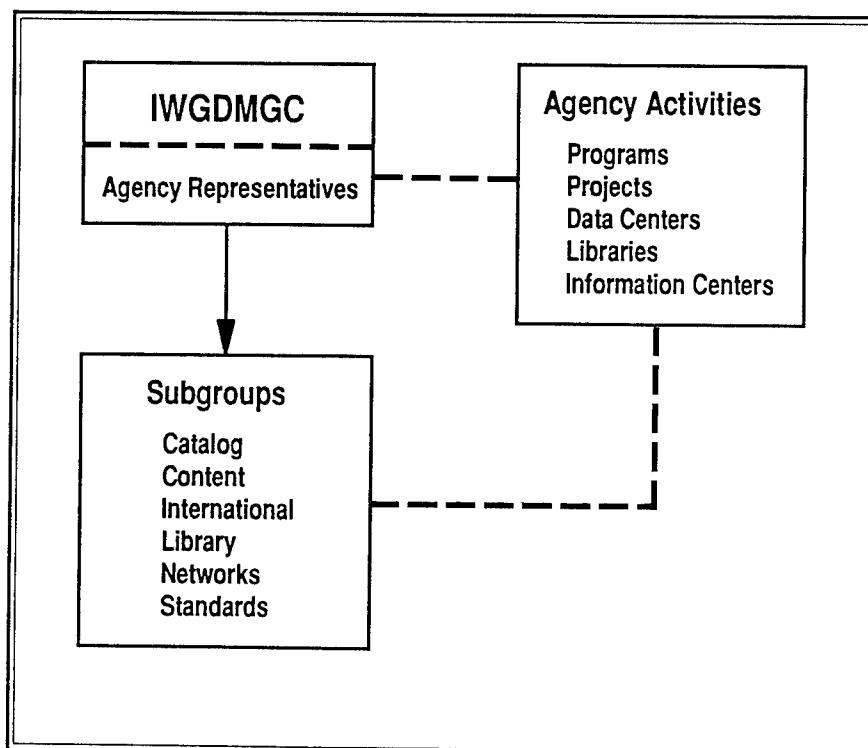
example, U.S. data policy requires that there should be at least one explicitly designated archive for each and every global change data parameter and that a clearinghouse should be established to prevent the purging and loss of important data sets. This plan provides for these national capabilities as well as others specified by the data policy statements (see Chapters 4 and 5).

As currently established by the CENR, guidance, review, and oversight of the USGCRP is provided by the NAS Board on Global Change and other relevant committees of the Academy. The Committee on Geophysical and Environmental Data, working with the Board on Global Change, serves this function for the data and information management component of the USGCRP.

## GCDIS Operations

**Figure 2. GCDIS Implementation Organization**

The GCDIS is an evolving distributed data and information management system that will initially consist of interconnected, loosely organized agency data and information centers. As the GCDIS evolves, the system will improve its connectivity, performance, and responsiveness to user requirements.



No one agency is in charge of the GCDIS. The GCDIS operates by participating agencies providing data content and data access services that have been identified, recommended, and coordinated through the IWGDMGC. The IWGDMGC is responsible for GCDIS implementation. This responsibility is carried out by IWGDMGC agency contacts, subgroups of the IWGDMGC, and the IWGDMGC Executive Secretary. The Associate Scientist for Data also provides coordination with the IWGDMGC in GCDIS requirements and implementation activities. Figure 2 shows the GCDIS implementation organization, and the organization that will be in place during initial operations of the GCDIS. Descriptions of specific data and information centers begin on page 83.

Users will be able to interact directly with the GCDIS at various levels of sophistication, ranging from a simple phone call to a data center to online search, browse, and network data set transfers. Initially, a user services group or contact at agency data centers, information centers, and libraries will provide assistance, guidance, and information. The extent of user services will vary within and between agencies, and will be improved in response to user needs and current technology.

For GCDIS-related content and access matters at the interagency level, users will also be able to contact the appropriate IWGDMGC agency representative, subgroup chairs, and agency contact members, or the IWGDMGC Executive Secretary concerning GCDIS problems, information, and comments. In addition, there will be a bulletin board system for users to get the latest information about the GCDIS, make suggestions, and request special interagency products and services.

## GCDIS Coordination

The IWGDMGC, through agency representatives, maintains the GCDIS implementation plan and coordinates data and information implementation activities of participating agencies. This coordination includes tracking the progress of individual agency implementation efforts—each managed within the agency—against the overall implementation plan. The IWGDMGC identifies problems and recommends actions to facilitate implementation.

The USGCRP has been organized around four interconnected streams of activity: observations and data management, integrative modeling and prediction, process research, and assessments. These activity streams, coupled with a supporting stream focused on communicating the findings to the public and to students, continue to provide a sound basis for carrying out an encompassing scientific program and are suited to strengthening the coupling to evaluation, assessment, and policy analysis. This functional architecture is depicted in Figure 1, which illustrates the

interconnections among these streams of activity and the growing links to policy development.

Within this structure, the participating USGCRP agencies continue to maintain control of their budgets. The CENR, through the Subcommittee on Global Change Research (SGCR), provides the interagency coordination, implementation, management, and oversight, including the interagency review of programmatic needs and budgets.

This coordination mechanism will facilitate interagency cooperation and will help ensure completeness of coverage of USGCRP high-priority needs, including links to crucial data and information from outside the USGCRP programs. To help this effort, full-time, collocated associate scientists for each of the areas (observations and data management, process research, integrative modeling and prediction, and assessments) have been assigned to the Office of the USGCRP.

IWGDMGC GCDIS coordination will focus on two primary areas: data content and data access. Data content issues address data and information content, metadata, documentation, user coordination and related areas (Chapter 4). A Data Content Subgroup will be responsible for coordinating these aspects of the GCDIS. Data access issues address how users gain access to the data and information of the GCDIS (Chapter 5). Data access issues will be coordinated by the combined efforts of the Catalog, Networks, Standards, and Libraries Subgroups.

Successful implementation of the GCDIS includes appropriate use of standards for both the enabling technology and applications-specific content. Wherever possible, there will be full coordination with existing standards bodies, and implementation of existing standards will be used. A Standards Subgroup will work on both data content and access issues and will recommend appropriate standards for adoption by the IWGDMGC.

Direct coordination among agencies is essential for achieving program goals. For example, those agencies upgrading or developing their own data and information systems as their components of the GCDIS will coordinate the technical details of the design and implementation of their systems within the framework provided by this implementation plan. Agencies are encouraged to invite other agencies engaged in parallel efforts to participate, where practical, in specific design choices or implementation decisions to help realize the interoperability required for the GCDIS.

Each agency will attempt to incorporate GCDIS recommendations into its own data and information system implementation whenever possible, given cost and schedule constraints. When an agency must delay or forgo implementation of a key GCDIS feature, it will notify the IWGDMGC as soon as possible.

## External Coordination

Liaison will also be maintained with government organizations at the State, Federal, and international levels, including some nongovernmental organizations. These liaisons will be maintained through the IWGDMGC, often through cross-representation by IWGDMGC member agencies in these external organizations. The IWGDMGC will conduct a thorough review of such organizations and will establish formal links with those organizations determined to have significant related purposes and activities.

It should be noted that versions of the U.S. global change data policy have been formally adopted, in whole or in part, by several international organizations. While not binding on the member countries, this represents an initial step toward establishing an international norm and the facilitation of GCDIS international linkages and coordination. Representative international organizations whose work has already been identified as being important to the GCDIS are identified in Chapter 7.

In addition, there are a number of other programs outside the USGCRP, evolving in parallel with the GCDIS, that can make important contributions to the broad information needs of researchers, policy analysts, and others attempting to understand the many aspects of our environment. Since users are best served through seamless access, the GCDIS will adopt or support interfaces that accommodate the information access approaches being developed in such overlapping national and international initiatives as the U.S. National Information Infrastructure, which includes the Government Information Locator Service and the National Spatial Data Infrastructure.

## Policy Coordination

There are aspects of data and information management that impinge on policies beyond the immediate USGCRP, but are addressed within the Federal structure for information and technology policy coordination. Examples include private enterprise—government interactions, the HPCC initiative, scientific and technical information policy, education initiatives, and the National Information Infrastructure (NII). In these areas, the IWGDMGC will coordinate policy issues through the structures established in the Federal Coordinating Council for Science, Engineering, and Technology.

As one example of such policy issues, the increase in public access capability creates an issue involving the confidentiality of data and information. Social science data, at the finest level of analysis, deal with individuals. Most social science researchers

protect the identity of individuals, and even groups, as a fundamental part of their research. Even nations restrict the availability of social and economic data and information or distort it substantially for political purposes. Mere aggregation will not necessarily solve the problem, since individual-level data may be needed for integrated assessments. USGCRP agencies and departments may need congressional action to provide the protection required for the confidential elements of the GCDIS.

This problem is not limited to the social sciences. Biological data and information, such as forest inventory data and agricultural statistical data, sometimes have economic and social (privacy) issues associated with their use. In the process of associating field data with pixels in satellite images, for example, sensitive location and ownership information may be revealed. The Federal mechanisms to protect these data, while making them available for USGCRP analyses, do not yet exist. They must be developed both technically and politically as a priority item. This will require action by the executive and legislative branches of the Government and will probably be tested in the judicial. Until adequate safeguards are in place, both biological and socioeconomic data will be limited to publishable data and information and thus compromise the ability of the USGCRP to meet its objectives.

## User Coordination

The GCDIS will be structured to make certain that users can provide their insights and opinions at all levels of the system. At the level of the IWGDMGC, such input will come from multiple sources. These include

- The liaison member on the IWGDMGC from the Committee on Geophysical and Environmental Data (CGED) and the Board on Global Change of the National Academy of Sciences (NAS),
- A continuing series of forum reviews by the CGED,
- Participation in IWGDMGC subgroup-level activities by user groups not represented by the CGED, such as library, education, and State government associations,
- Regular user surveys based on the GCDIS measures of success (see page 6),
- The CENR and the working groups of the USGCRP, each of which has its own user advisory processes, and
- The agency members of the IWGDMGC.

At the agency level of the GCDIS, the IWGDMGC will continue to establish general policies for user coordination. The program plan and this implementation plan are examples of this coordination process. Within this process, agencies (or their designees) providing GCDIS access will have their own user-advisory structure.

These agency structures will remain the responsibility of the particular agency and will include data center advisory committees, project and program advisory committees, and a multitude of types of interactions with individuals—both those who supply and those who use GCDIS data and information.

Although the GCDIS initially is designed to meet the critical needs of the global change research program, new categories of users will be accessing the system as it evolves. Several factors will contribute to broadening interest in the GCDIS. For example, the links to contributing data and information from the natural resources and human dimensions areas; the increased focus on data and information supporting global change assessments; and the availability of tools to assist users in identifying, accessing, and combining data and information in meaningful ways will open doors to new classes of users. These will include users such as students below college level, policymakers, and environmental interest groups who find the current system too limited in scope or too difficult to use.

These new users will bring fresh viewpoints that will be invaluable to the continued development and measured success of the GCDIS and that will require specific outreach programs to obtain their input. Each of the participating agencies has the responsibility to reach out to the users of their data and information because the agencies can more readily identify their own users and have more opportunity for informal and formal contact with them.

### 3. Users and the GCDIS

Several different types of users will require data and information contained in or pointed to by the GCDIS. Each will need different means of information identification, access, and documentation, and each will need different information products to suit its applications.

User requirements will fall into three broad categories. The first major category of users will be the general research community, which will primarily need access to raw and lightly processed data for basic studies. This group will include discipline-oriented researchers, graduate students, and educators requiring data and information for advanced college courses. This group will require sophisticated browse capabilities, and will have access to the more sophisticated computers, interfaces, and networks. The data volume for this group will be the largest of the three general user communities, but the information content per datum will also be the lowest.

The second major user group comprises those needing summary or multidisciplinary integrated data sets that have been assembled by researchers or data centers to address a particular global change problem. These higher-level integrated data sets are assembled from primary discipline-oriented data. Examples of such products include long-term climate records (carefully quality assured) used to estimate historic climate changes and integrated geologic, geographic, oceanographic, ecological, and meteorological data sets used to predict the effects of potential sea level rising. Access to such data sets will require less browse capabilities and will be used by researchers and others with less data-handling capability. These data products will also be used by those investigators pursuing integrated assessment activities or teaching environmental courses.

The third group of users will be those interested in high-level information products that directly address global change issues in ways helpful in decision-making processes for anticipating and adapting to environmental changes. These products will be rich in information and might include, for example, summaries of expected changes of agricultural crop ranges due to temperature and moisture changes. Such products have the highest information density and could also be used by those in private enterprise and government concerned with planning for changes in the projected economic basis for a region's industry. That such products are directly linked to recognizable environmental conditions may make them the most useful data products for the general public and primary or secondary school use. Access to these products may be through more traditional means, such as responding to paper

announcements or phoning data centers directly. Their actual production media may often be paper based.

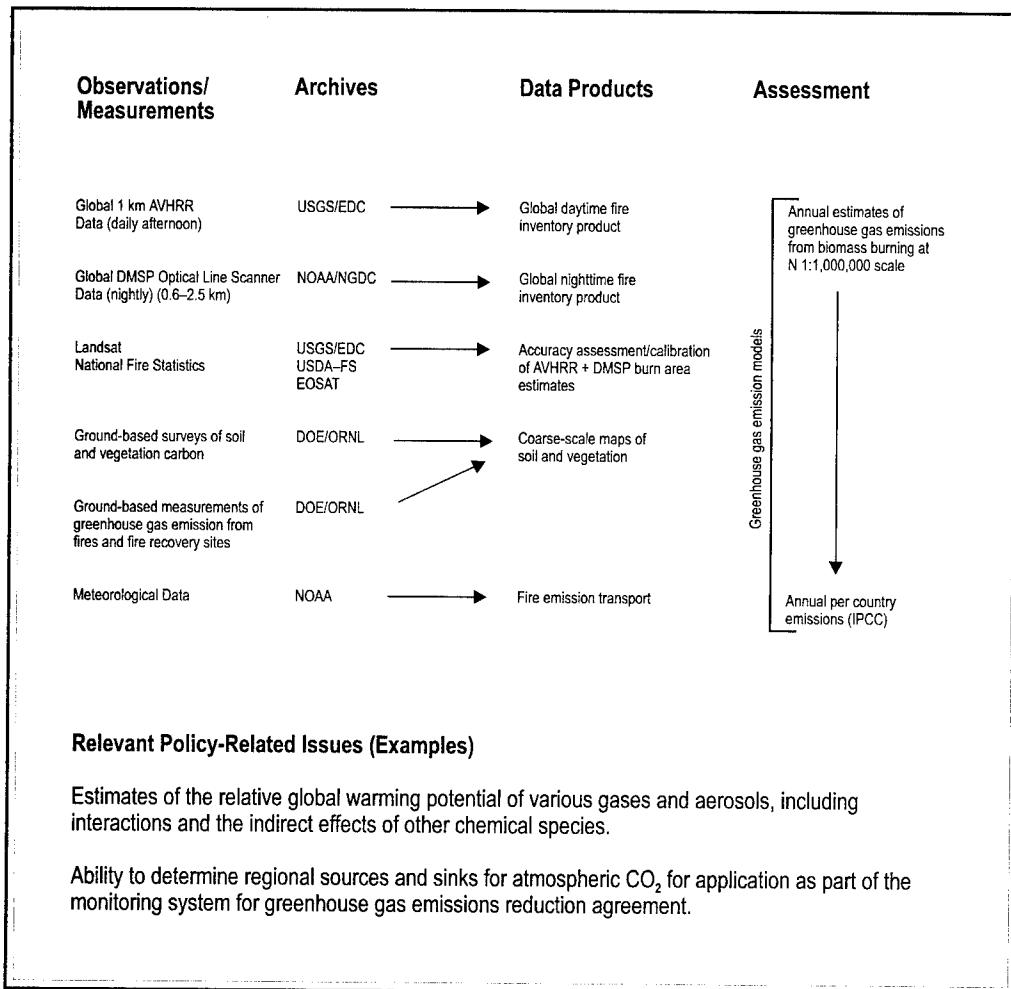
Users in the second and third categories will need specially tailored GCDIS directory capabilities for locating the data and information applicable to their initial needs without being confronted by the confusing (to them) amounts of data and information available through the GCDIS system.

## Diversity of the Users

The complexity of global change issues is reflected in the diversity of the potential users of the GCDIS. Since the GCDIS will be responsive to users of the system, it is important that there be a systematic approach to characterizing users in order to understand their needs better. The design and implementation of a system like the GCDIS depend on the nature of the functions the users are performing in their interaction with the system. The access components of the system will depend on the sophistication of the users or their intermediaries, and the contents will be driven primarily by the types of questions the users need to answer within their own particular problem domains.

For example, the following user scenario (and Figure 3) illustrates that an individual researcher doing a scientific assessment can take advantage of the internal connectivity provided by the GCDIS.

International negotiators of the United Nations require the answers to many policy-relevant questions on greenhouse gas (GHG) emissions, including, for example, the importance of global, regional, and national biomass burning to the emissions of GHGs. In this scenario, the CENR SGCR assigns the science assessment of the importance of biomass burning to EPA, which designates a manager and forms an interagency, international team that defines the scope of the investigation to answer two questions (see the bottom of Figure 3) on the current status of emissions and their sensitivities to natural and managed changes, and future emissions for various managed scenarios.



**Figure 3. Greenhouse Gas Emissions from Biomass-Burning Scenario**

The science assessors agree to attack the questions by producing a science assessment report that will be produced in sequence (moving from right to left in the diagram) from global or regional modeling products, data products, archives, observations, measurements, and process studies. The science assessors specify the accuracy, precision, representability, and comparability characteristics of the modeling and data products to permit the integration of the products as the basis of the science assessment.

The data and modeling products may be either pure data bases from the archives, transformed data bases, or data bases synthesized by a combination of various

observed data, process results, and modeling results. In this example, several archives are combined to form synthesized data products (e.g., Landsat data and National Fire Statistics to give an accuracy assessment or calibration of combined Advanced Very High Resolution Radiometer (AVHRR) and Defense Meteorological Satellite Program (DMSP) burn area estimates).

The role of the GCDIS is critical in forming data products. The archives shown must be part of GCDIS for ready access, comparability problems must be solved to allow merging of data bases, and data and modeling products must be placed in accessible archives.

## **The Problem Domain of the Users**

At the highest level, the problems addressed by the GCDIS user community can be organized in parallel with the four working groups associated with the USGCRP: observations and data management, process research, integrative modeling and prediction, and assessments. The USGCRP, through the four working groups, defines the problem domain for the GCDIS. The process for implementing this problem definition is outlined in Chapter 4 of this plan.

As a problem domain example, there are special needs that come from considering the problem domain of a user interested in process research. A wide variety of processes exist, ranging from basic meteorological processes through ecological processes to social and economic processes. Further, the problem domains in process research interact strongly with data access as well as data content issues.

One aspect of meteorological research aimed at understanding climate-related processes is the deployment of field programs. There are two aspects of the GCDIS user community involved in field programs. On one hand this community relies on the GCDIS to preserve and distribute its data and information, and on the other it also requires data and information from the GCDIS to support both the operation of the experiment and interpretation of its results. In general, the needs of this community for data and information preservation and analysis support are well known. The support of field programs, however, can generate a requirement for the timeliness of data delivery that can have an effect on the GCDIS considerably different from an archive-oriented user strategy. In particular, the nature of quality control procedures may change, and the necessity for standard data exchange protocols becomes even more important.

## **The Uses of the Data and Information by the User**

Approaching the user community according to the way the user will use the information is important. The use of the data and information may depend on the skills or the sophistication of the user, recognizing two important factors: the level of expertise of the user in locating and accessing data and information, and the user's level of understanding of their use.

There are three levels of individual interaction with respect to data use: interacting directly, which leads to personal use of data or information; acting as an agent for someone requiring data or information held within the GCDIS; and acting as an intermediary who not only accesses the data or information, but also performs a value-added function upon the data for a user who does not choose to, or who cannot either interact directly with the system or perform the analysis.

By addressing the sophistication of the user, those implementing the GCDIS will make it possible to decide explicitly whether a user group should be served as though it were going to interact directly with the GCDIS or whether it would be more appropriate for the user group to interact through agents or intermediaries. In this fashion, design criteria will be developed for software to ease user interaction with the system.

Although there will be activities within the GCDIS that provide value-added products, a substantial portion of the products of the GCDIS will be generated by entities external to the GCDIS. These entities, acting as intermediaries, will range from international scientific assessment groups to policy analysts at the State and Federal level, to authors of monographs and textbooks, and to the press. These intermediaries are the trained analysts, educators, and information specialists who have the expertise necessary to convert data into information for the myriad of users that have a need for global change information, but not necessarily raw data.

These external intermediaries include the four major working groups of the USGCRP. In this context, it should be noted that the GCDIS not only supports the working groups as their activities generate products for external users not in immediate contact with the GCDIS, but it also facilitates the interchange of data, information, and analytic results among the working groups. A major implementation goal of the program will be to facilitate this interaction and to promote the achievement of the goals of the USGCRP.

The use of an intermediary by a high-level decision maker is illustrated in the following example.

To help formulate a position on a bill before Congress, a Representative from Louisiana needs the best possible information about this question:

Will sea level rise caused by global warming inundate parts of my congressional district in coastal Louisiana? If so, how can the adverse effects on my constituents be alleviated?

The Representative asks a staff member to provide this information before the vote on the bill. The vote is scheduled to occur in a few days.

Members of the staff have been made aware of the capabilities of the GCDIS by IWGDMGC-coordinated outreach efforts. If staff members have access to the Internet, they will electronically access the GCDIS directly and find the required information. If they need an intermediary to obtain GCDIS access, they will call a group such as the Congressional Research Service for help.

Since the GCDIS has experts and agency systems linked by interagency cooperation and networks, the staff quickly obtains a comprehensive view of the data and information available about the Representative's question. The staff finds the required information and reports the next day that some global change model predictions do in fact predict a large enough sea level rise to inundate certain parts of low-lying coastal areas (such as Louisiana) that contain substantial population, potentially serious environmental pollution problems (such as abandoned hazardous waste sites), and so forth. The staff also reports that continuing efforts are being made to evaluate the accuracy of the sea level rise predictions, assess their multiple effects, and explore possible mitigation measures and their costs.

From this example several points can be made. First is the importance of a problem statement. The staff member or the information specialist in the Congressional Research Service needed to have a clearly articulated question to interact with the system—the question provided by the Representative. Next, the Representative did not actually interact with GCDIS. An intermediary was used to interact with the system and obtain data for someone else. It is important that how the system was used did not depend on the fact that a Representative asked the question. Anyone acting through the intermediary could have gotten the same answer. The Representative may have been able to investigate the matter, but, as in the case of many information-related activities, there exists an existing information infrastructure that is ready to use the GCDIS to meet its function of providing information.

## Organizational and Group Affiliation of Users

While it could be argued that the immediate user of GCDIS is always ultimately an individual, individuals acting in groups or as part of organizations can take on very specialized roles. These roles can affect the range of activities that need to be supported by the GCDIS, and the group or organization may place requirements on the system that might not normally be part of an individual's approach to the GCDIS. Group or organization requirements can be particularly important when delivering data to the GCDIS.

**Table 2. User Types**

	Single User	Small Groups	Major Programs and Communities
<b>Observations and Data Management</b>	Researcher validating a GCM University student	Research group working on long-term ecosystem variability (UNH)	USGCRP research areas ISSCP Educational community International global satellite systems
<b>Process Research</b>	Individual researcher improving a model	Research group working on El Niño and the equatorial Pacific thermostat	USGCRP research areas GEWEX, FIRE
<b>Integrative Modeling and Prediction</b>	Agricultural researcher analyzing effects of climate change on crops	Group comparing the output differences of various GCMs	USGCRP research areas CHAMMP
<b>Assessments</b>	Policy analyst (State or Federal)	IPCC assessment groups	USGCRP research areas The press Educational community

The organization of the user group is important because the nature of the GCDIS implementation can be very different if the user is an individual, part of a user group, or a member of a large team of researchers involved in a major field program. Table 2 illustrates the different kinds of users that will interact with the GCDIS. The column labeled *single user* identifies individuals who might interact with the GCDIS and what they might try to accomplish with that interaction. The columns referring to aggregations of individuals give examples of real organizations that are likely to be

represented in their interaction by an agent (usually the case of small groups) or that may be acting as an intermediary with the GCDIS.

## **Libraries and Information Centers as Intermediaries and Facilitators**

Libraries and information centers are major gateways to data and information resources. Librarians and information specialists are experienced in organizing, categorizing, conceptualizing, retrieving, and disseminating information. These skills and knowledge are critical to many GCDIS users who need to navigate the vast amounts of printed and electronic GCDIS data and information.

As the USGCRP evolves, there will be increasingly more printed and electronic material about global change issues and research. Much of this information will appear in books, reports, and journal articles scattered throughout the world's literature and in data bases accessible over electronic highways. Some research information will describe the state of understanding of global change and some will suggest further observations needed to test proposed models. Policymakers, students, and the informed public will find such interpretative information necessary.

Integrated data and information, at various levels of detail, will be most useful for such users. In addition to integrated data and information, research users may need digital and analog data directly. Common to all types of users is the potential value of the librarian or information specialist—not only to help users navigate the GCDIS, but also to synthesize the available data and information into the particular products and knowledge needed by users.

In a number of the participating GCDIS agencies, management of data is handled by major data centers or other organizations. Similarly, the management of much of the agencies' published information is handled through centralized Scientific and Technical Information (STI) programs. These STI programs comprehensively aid in collecting literature produced by their agencies, their contractors, and through their agencies' cooperative international and other research and development efforts. This literature is compiled into bibliographic data bases for searching and retrieving. The central agency STI programs also maintain systems that provide archiving and full text retrieving of such material. The USGCRP will take full advantage of these STI services in the Federal agencies.

The GCDIS will encourage efforts by libraries and information centers—of all sizes and types—to collect and make available the needed global change research data and information. To help such institutions in their roles as intermediaries and facilitators, special library and information center capabilities will be developed when necessary, and the GCDIS will take a proactive approach in soliciting their continuing feedback, comments, and recommendations.

### User Scenario

An example of a library user of the GCDIS is a graduate student working with interbasin water transfer. By using the capabilities of the university's library system, the researcher accesses the GCDIS to obtain both data and information on rainfall, water supply, water demand projections, and details from documented proposals for alleviation of the regional water shortages. The researcher browses various water deficit scenarios depicted on regional contour maps. The sensitivity factors, such as rainfall, are accessible for browsing or manipulation. Water management plans are searched for interbasin proposals that were implemented in areas having comparable population sizes. The researcher scans the management plans for text concerning project cost and environmental impacts, and then uses the results of the search not only to determine both the data and information available on this subject, but also to guide the future of the research. Throughout this process, the librarian or the information specialist is available as the human interface, or intermediary, easing the task of data system navigation.

### Implications of the User Attributes on the GCDIS

A critical step in implementing the GCDIS is the systematic examination of users by the Content and Access Subgroups to establish priorities for the content and access implementation strategies. In doing this it is very likely that several things will happen.

With respect to content, interaction with users about addressing their problems will lead to the identification of particular data sets outside the USGCRP. For example, the Carbon Dioxide Information Analysis Center (CDIAC) at Oak Ridge National Laboratory estimates that more than half of the data in its data center are not of U.S. origin. Also, the Environmental Protection Agency (EPA) has indicated that a substantial fraction of its data is generated by and held by the States. These examples show that the satisfaction of user demands will lead to an expanding network of data and information that will necessitate strong and reciprocal interaction with both international partners and State agencies. This implies that an evolutionary growth of the GCDIS will continue for the life of the program, and that many of the holdings will be the result of interactions with users beyond the USGCRP.

With respect to user access, the data and information must be delivered to users in a fashion consistent with their use of the data and information. Issues of connectivity of results, quality of metadata, and the timeliness of delivery of the data and information as indicated in the scenarios will be critical to successful implementation. As the use of GCDIS data changes, access and content also will change. This requires an active user feedback mechanism that will be implemented at each user contact.

Finally, additional complexity for GCDIS implementation is derived from GCDIS development originating from a collection of individual agency systems—each with its own existing user community. This is especially true for data sets from outside the USGCRP focused programs. The GCDIS users are effectively superimposed on the existing user community, whose composition, attributes, and needs have driven fundamental data base design, composition, and access decisions that may or may not be appropriate for GCDIS use. Individual agency implementation plans will move toward reconciling these differences by adopting common standards and guidelines.

## User Support

The GCDIS will identify a point of contact for user questions about each data and information source. For example, the GCDIS will identify a user support source for each directory or data source to which the GCDIS provides a pointer and also for each data set and information collection in the Global Change Master Directory (GCMD) itself. The minimal requirements for the GCDIS to provide pointers to particular data sets will be the continuing availability and the identification of a continuing source of user support for that data set.

The GCDIS will provide support to the users through a variety of mechanisms at whatever level the users interact with the system—as an individual for personal use, as an agent, or as an intermediary. In most cases, this support will be provided by the agency or other entity that is the repository of the data or information in question. Service will be provided with the distributed support of libraries, information centers, data centers, and individual sources. To provide for the diversity of users, the service providers will employ a wide range of information technology, service, and search tools and mechanisms, ranging from manual, print-based resource guides to the most sophisticated computer-based expert systems. The specific types of user support services for the GCDIS will be developed in stages based on both actual user needs and on what is technologically feasible and affordable.

A bulletin board system will provide centralized support to users, such as the latest information about the GCDIS. It will be a means for users to share experiences in using the GCDIS as well as a mechanism for users to provide input on problems and suggestions for change—including requests for special interagency products and services. Agencies participating in the IWGDMGC will evaluate user inputs received through the GCDIS bulletin board in their more formal agency reports on customer satisfaction consistent with Executive Order 12862, *Setting Customer Service Standards*. A major portion of the bulletin board will be operated by the Global Change Research Information Office (GCRIO).

The USGCRP established the GCRCIO in 1993 within the Consortium for International Earth Science Network (CIESIN). The purpose of the GCRCIO is to disseminate to foreign and U.S. individuals and organizations scientific research information available in the United States that would be useful in preventing, mitigating, or adapting to the effects of global change. Worldwide sources of data and information are available to the GCRCIO to satisfy requests about global change topics. These sources are being documented and made available over the Internet for use by GCRCIO staff and researchers worldwide. Most are available at no charge to the GCRCIO user.

The GCRCIO uses information services that already exist. A primary source of information is the GCDIS. Users are pointed to the GCDIS as one of the sources of information available from the GCRCIO. Documents and online information about the USGCRP are also referenced. The GCRCIO functions as a major access facility for worldwide users to the data and information available from the USGCRP.

Gerald S. Barton, Director, GCRCIO  
1825 K Street NW, Suite 805  
Washington, DC 20006

Phone: 202-775-6607  
Fax: 202-775-6622  
Internet: [gbarton@gcrio.org](mailto:gbarton@gcrio.org)

## 4. Content System Implementation

This chapter discusses the responsibilities of the Content Subgroup relative to GCDIS implementation. Included are a conceptual view of GCDIS content; scope of the Content Subgroup responsibilities; coordination activities necessary to facilitate interagency cooperation, interactions with State and local governments and international organizations; discussions of the content functional areas; and future considerations.

### **Conceptual View of GCDIS Content**

The USGCRP requires data and information of many types and from many sources to achieve the understanding needed both for developing a predictive capability for global changes and to assess the effects of such changes. An objective of this implementation plan is to provide an interagency framework for the coordination of the Federal agency data and information activities related to global change.

This implementation plan recognizes that data and information management activities designed for a particular agency program are most responsive to the community of users within that program. In responding to mission objectives, agencies have used diverse data and information management approaches. It is the thrust of this interagency global change data and information management effort to help coordinate the multiple individual agency programs related to global change and provide the necessary broad interdisciplinary capability needed by the USGCRP. This coordination will not only lead to the availability of more comprehensive data and information, but also help both to reduce redundancies and increase the ability to use data and information from disparate sources.

A major focus of the GCDIS implementation plan is to develop procedures and criteria for establishing priorities for some of the key data and information sets needed to meet the objectives of the USGCRP. The establishment of priorities will be implemented by actively involving the other organizational elements of the USGCRP, the NAS, and the research and user communities. In following sections, three levels of service, corresponding to three levels of priority, are provided as guidelines for the agencies in implementing specific functions. These levels of service are summarized in Table 3.

**Table 3. Data Content Functionality**

Priority	User Coordination	Search and Locate	Documentation	Quality Indicators
1	Users involved in all facets	All sources	In GCDIS. Full set of functions available for use with full confidence	Evaluated by peer review
2	Users involved in priorities, assembly, documentation, assessments, etc.	All U.S.	In GCDIS. Full set of functions with detailed description	Evaluated by an agency
3	Users involved in advisory functions	All agencies	Available through GCDIS with summary	Checked by an agency
Not rated		As submitted	Available functions defined in GCDIS	As available

## Scope of Content Subgroup

The Content Subgroup will provide the links between the data and information needed to answer key USGCRP questions and the activities required to provide the data and information. The subgroup will identify the data and information contributions from each agency and ensure that they are available through the GCDIS to the other agencies, users, and working groups of the USGCRP. To enhance the intercomparability of data and information, the Contents Subgroup will work with the Standards Subgroup to provide appropriate FIPS and voluntary standards and guidelines to be adopted by agencies in their implementation plans relevant to the delivery of information about data and information holdings that include descriptions, quality assessments, supporting information, and guidance and aids for locating and obtaining the data and information. To expedite implementation of these criteria and guidelines, mechanisms will be developed for agencies to share experience and expertise.

A particular activity of the GCDIS implementation will include coordinating activities needed to produce special data and information products needed by the users of global change data and information. These will include value-added data sets that require data from multiple agencies. For example, assessment studies will require the synthesis of models from individual parts of the global change system. Process studies data sets that result from the synthesis of data from different platforms will also be needed by researchers in other agencies or their programs.

Information summarizing the global environment and its change will be produced by the USGCRP; these results will have high-priority availability. Many of the summary products will be in text, graph, and table formats rather than in numeric.

## Coordination of Content Implementation

Working under the framework of the IWGDMGC, the Content Subgroup will provide programmatic focus to develop and coordinate this framework. User oversight will be provided through formal mechanisms with the NAS and other advisory bodies. The subgroup will use ad hoc panels to address issues; to develop guidelines, priorities, and processes; and to review other data and information topics.

The Content Subgroup, in coordination with other elements of the USGCRP, will identify and address those issues that affect the content of the GCDIS. This will include data from Federal agencies, State and local governments, and international organizations.

The ad hoc panels will comprise experts identified by the agencies, and will be called on to respond to many of the content issues. The panels will assist in the coordination of identifying and providing data holdings and special products needed to respond to the user needs—for example, interagency efforts such as the Global Energy and Water Cycle Experiment (GEWEX).

## User Coordination

The user is the central focus in the GCDIS implementation plan. Users both contribute and draw on the system's data and information resources. Thus, the research and other user communities must be actively involved in all facets of the system—design, implementation, and maintenance. Specific aspects of user coordination are:

- Identify the parameters that are key to global change research, in building a priority list of global change research parameters required to achieve the recommended milestones.
- Assemble and document priority data sets, and develop and recommend value-added data and information products for subsequent analysis (e.g., browse products, literature, citations, bibliographies).
- Identify priority data and information resources to be accessed, acquired, or developed.
- Participate in advisory committees toward the development and enhancement of policies and procedures for archiving and purging data.

- Identify international partners to share global change data and information for mutual benefit.
- Assist with the development of cross-training opportunities (e.g., workshops, conferences, programs, feedback mechanisms) for computer and information scientists and others to interact with global change researchers, including university programs for the purpose of reviewing and identifying problems and improvements.
- Provide feedback in determining problems, gaps, omissions, or errors in the coverage for priority data sets and recommended remedies.
- Assist with the validation of selected data and information products through such processes as peer reviews and assessments after actual use.
- Deliver selected derived data and information, with documentation, from process studies, predictive models, or assessment activities to the appropriate data center and/or user.
- Assist with participation in projects to assemble retrospective data sets and data rescue projects when needed to preserve crucial data that are at risk.
- Assist with building upon existing digital and nondigital data and information resources to improve access to high-quality global change data and information by integrating appropriate activities or participant data centers, archives, libraries, and other information-disseminating organizations, and providing products in appropriate media, depending on the particular user's need.
- Help solicit active involvement of the national and international research and user communities in reviewing the effectiveness of the GCDIS, in facilitating access to and use of global change data and information, and in articulating the changing needs of the GCDIS users.

Level of Service	Level 1	Users involved in all facets listed here.
	Level 2	Users involved in defining data priorities, data set assembly, definition of special product needs, and the definition of other content related issues.
	Level 3	User feedback and advisory group participation.

### **Coordination of Research Data Input**

The Content Subgroup, in coordination with other elements of the USGCRP, will investigate various mechanisms to encourage submission of data sets to data centers

for broader use. This activity will focus on those data sets that are produced in USGCRP research projects that have resulted in peer-reviewed publications. Possible avenues include contractual as well as professional incentives. Discussions with professional societies will be initiated to investigate journal status for the preparation of data sets for use in the broader community. These discussions will address both the inclusion of data set articles and citations for the use of data sets in other research publications.

### **State, Regional, and Local Government Coordination**

State, regional, and local governments generate data and information that may be necessary for climate and global change research. Dialogues with these government entities will be developed to ensure that their data and information are made available to the GCDIS and that the GCDIS data and information are available to these government organizations, in return. Examples of relevant data are inventories of biogenic and anthropogenic emissions, land use and comprehensive plans, air and water quality measurements, biodiversity data, and inventories of living resources.

Mechanisms and links to State, regional, and local governments must be developed to transfer priority data, metadata, documentation, and related software to appropriate local or national archives. Procedures will be developed in the mid-1990s to begin making these data and information available and to integrate these data and information into the existing archive systems. These procedures will include potential funding mechanisms to assist States with the transfer of nonfederally funded data. The procedures also will include recommendations for transferring data collected for federally funded projects to an archive accessible by the GCDIS. Examples of existing federally funded programs are the EPA State Data Management Program and the National Oceanic and Atmospheric Administration (NOAA) Coastal Zone Management Program. The USGCRP can use these existing funding mechanisms as models for future funding to State and local governments as part of the USGCRP budget.

### **International Coordination**

Most high-priority data sets are global, and, since the extant global data sets are generally compilations of older data that have disparate spatial scales and represent regions at different times, new global data sets are needed. In many cases, it is most practical to recompile new global data sets from newer, better documented, higher resolution data directly from the scientific groups in larger countries or in regions. Some pilot data projects recently finished at NOAA's National Geophysical Data Center (NGDC) in Boulder, Colorado, indicate that this approach is practical. It involves direct collaboration between data groups in the United States (National Data

Centers [NDCs] or World Data Centers [WDCs], for example) with scientists and institutions in other countries, particularly in developing countries, since it may be assumed that modern data will be more easily exchanged with developed countries.

Some U.S. global change scientists and institutions already know of such modern data in developing countries. Otherwise, it may be necessary to explore the existence of such data through the International Geosphere-Biosphere Programme (IGBP) committees of these countries or through national institutions. The WDC system also may be of help in searching out such modern data sets.

The Content Subgroup will identify disciplinary regional and global data and information that are known to exist but that are not available within the United States. The subgroup also will identify potential priority disciplinary regional and global data and information for which a source is not known.

Rescue of data needed for global change research but in danger of deteriorating or no longer being archived is another activity that will be actively pursued while the data and information in question are still accessible. The process of rescuing such data will include selective conversion to digital format. Another form of data rescue involves digitizing older tabular data, especially when these data represent important time series, to provide these data in maximally useful form for computer-assisted analyses.

Successful data rescue and digitization projects have been conducted with colleagues in Russia and China by units of WDC-A, at low cost. These collaborative projects can also be of benefit to scientists in the respective countries by providing access to technical assistance and data products. For example, receiving copies of their own data back in digital form (on compact disk-read only memory [CD-ROM] devices, for example) may make their data more useful for national studies.

## **Data and Information Identification, Acquisition, and Priorities**

Identifying and setting priorities for data and information sets necessary for global change research are iterative activities. These are the first steps in a process that involves including data and information sets in an archive, collecting (or creating) a data and information set if it does not already exist, managing data and information, and distributing data and information to GCDIS users. A recognition of the critical importance of data and information management, including necessary resource requirements for each step, is key to making the best decision to include a new data and information set for archive and distribution.

The Content Subgroup, as a standing group, will make links from the data and information needed for key USGCRP research questions to the identification of,

provision of, and priorities for the necessary data within the GCDIS. This will be done in the context of the USGCRP framework, which provides overall direction for the collection of data and the use of information for supporting global change research.

To provide such direction and to coordinate the interagency activities necessary to meet the goals of the USGCRP, four parallel, but interconnected, streams of activity and working groups have been established. These are

- Documenting global change (Working Group 1: Observations and Data Management),
- Understanding key processes (Working Group 2: Process Research),
- Predicting global and regional environmental change (Working Group 3: Integrative Modeling and Prediction), and
- Assessing and synthesizing the state of scientific, technical, and economic knowledge and implications of global change (Working Group 4: Assessments).

All four streams not only provide data and information, but also have specific requirements that the data and information management program must address. These requirements are driven by questions that the USGCRP must answer. These questions, in turn, may be used to set priorities for data content in the GCDIS.

The four working groups will be asked to provide priority questions that require information from the GCDIS. Expert advice will be sought from the agencies and from outside groups, such as the NAS, to develop detailed lists of data and information needed to develop the information to answer the USGCRP questions. Priorities for data and information will be derived from priorities associated with the questions that generated the information request. These data and information priorities will then be used to develop priorities for data and information from archives external to the USGCRP and will be used by the Access Subgroup to develop the level of implementation required for data and information in the GCDIS. In addition to this primary method of setting priorities for data and information, external advice will be sought from a number of sources.

To that end, the Content Subgroup will survey existing data and information, such as the documentation of scientific reports, assessment reports by the NAS, and written reports by the international scientific community. In addition, the subgroup will survey written reports from time to time and will attend meetings on an informal, member-by-member basis in order to identify new data sets that the process just described may have missed. Review of new national and international program requirements as they become available will be coordinated by the Content Subgroup.

To determine the availability of needed data and information, the subgroup will survey such current holdings as are defined by existing catalogs in the GCDIS. In this manner, gaps in the needed data can be determined. The survey will also identify and describe library and data collections important for global change use, beginning with GCDIS agencies' collections, then other Federal collections, then non-Federal collections (e.g., universities', State and local governments', nongovernment organizations', international). This information will then be made widely available through the GCDIS directory.

The Content Subgroup will use the existing international framework for identifying international data and information that may be needed. Representative organizations include the Intergovernmental Panel on Climate Change (IPCC), which was established by the World Meteorological Organization (WMO), and the United Nations Environment Program (UNEP). The IPCC assesses climate change for policymakers every several years, focusing on immediate issues and consequences. The IPCC prepares in-depth summary information for policymakers so that realistic response strategies can be created to manage climate change issues.

Many of the requirements underlined in the IPCC documents are addressed by continuing international programs. For example, international projects are coordinated by the World Climate Research Programme (WCRP), sponsored by the WMO and the ICSU, and by the IGBP, sponsored by the ICSU. Both the WCRP and the IGBP have addressed questions of data set identification.

For example, an IGBP report<sup>1</sup> presents a summary of the scientific requirements for a 1-km data set. The identification of such a high-priority need by IGBP and other scientific groups led to a high priority for the collection of the data set. An international and interagency Global Land 1-km AVHRR Data Set Project is now underway. This is an unprecedented effort to collect a daily global data set for a 1.5-year baseline. The project includes the U.S. Geological Survey (USGS), NASA, and NOAA (in the United States) and the European Space Agency and Australia, along with several dozen high-resolution picture transmission stations as participating collection sites. The Committee on Earth Observations Satellites (CEOS) has endorsed this project, and asked its Working Group on Data to provide support needed for international coordination of project data management issues.

---

<sup>1</sup> IGBP Report No. 20, 1992. *Improved Global Data for Land Applications*.

## Assembly

Data and information set assembly, or acquisition, is a major function within the GCDIS. Assembly functions will be implemented in conjunction with and closely coordinated with the identification, location, and priority-setting functions described in this implementation plan. Assembly involves basic data and information set acquisition, along with related ancillary data, quality assurance, and data set documentation, by the responsible agency in order to provide data and information access and archiving functionality. For some high-priority data and information sets, reassembly may be necessary where the data and information sets must be reformatted or migrated to higher density media, such as CD-ROM or 8-mm tape, or combined with other data to develop value-added, multidisciplinary data bases.

The focused programs of the USGCRP are major sources of global change data and information. Resource requirements for the focused programs' associated data and information management functions have been included in the USGCRP planning. Other major sources are the many Federal and other programs with objectives primarily for purposes other than global change whose data and information are vital to the USGCRP. Examples include the daily satellite and in situ weather observations, information output from climate models, tree ring and ice core measurements, forest inventories, biological and ecological observations, cartographic data, stream flow records, bathythermograph data, fossil fuel statistics, demographic data, and soil maps.

Such data and information from these other programs are critical not only for global change research, modeling, and assessments, but also to the focused data and information program to fill gaps in coverage, tie together diverse data sets, and improve the quality and usefulness of the data and information by providing ground truth at selected points for calibration. The latter is specially important for the remotely sensed data required by the USGCRP. Some of these critically important data and information now are available to the USGCRP because they are intermingled with material subject to security, proprietary, or regulatory constraints, or are the property of entities outside the Federal Government, such as State and local governments, international sources, or individual researchers.

These other programs do not have global change as their primary theme. Therefore, the resources needed to make these data available to global change researchers (appropriate formats, products, and so forth) may not be available. A special effort is required to make the high-priority data and information from other programs available to GCDIS users in useful forms. The framework that will be used to accomplish this required aspect of GCDIS implementation is described in Appendix A on data capture. The IWGDMGC will coordinate with these other programs through appropriate interagency mechanisms to avoid duplication and to maximize efficiency.

## Search and Locate

A process of continuing search and review for global change data and information will be used to uncover data and information to be included in the GCDIS. Using the programmatic guidance available, the Content Subgroup will identify those data and information potentially available in Federal agencies. Using this and information about other volunteered data and information, the Content Subgroup will then determine what additional data and information are needed. The scope of additional searches will be guided by the priority level of the particular data or information set needed.

Level of search service	Level 1	All sources
	Level 2	All U.S. sources
	Level 3	All Federal sources
	Not rated	As made available

## Ancillary Data Streams and Metadata

The assembly function also will identify and provide ancillary data and metadata necessary for users to understand fully the characteristics and quality of the data and to be able to interpret results accurately. Important ancillary data include data-set-specific sensor calibration, sensor degradation rates and biases, geolocation data, and quality control information. Important metadata include calibration algorithms, sensor information, measurement procedures, correction factors, functions for derived values, experimental design and purpose, and so forth. It is important to point out in the metadata the weaknesses of the data set and both its appropriate and inappropriate applications.

## Data Set Documentation

Documentation (or metadata) is the bane of most research data sets. Unless developed by professionals, it often is inadequate, incomplete, and perhaps wrong. Important global change data sets must pass the 20-year test—can someone 20 years from now, not familiar with the data, fully understand and use the data set (perhaps for an analysis unanticipated at time of data collection) based solely on its printed and electronic documentation? This is a serious test, but one that must be passed for many of our data collections if long-term, global environmental programs are to be useful well into the future.

Complete documentation must do more than just describe the values represented in each field and the format information needed to read the media. It must fully

document the data set from all possible points of view. The extent of data documentation, as with quality assurance, will depend on the importance of any particular global change data set. At a minimum, data set documentation for those data sets deemed of highest priority should contain the following:

- *Identification of contributors.* Who are the principal investigators providing the data? It is important to identify original investigators so that they obtain credit for providing data to the archive and so that users can know the history and perhaps the quality of the data. No matter how much work the archivist does to prepare the data set for publication (quality assurance, reformatting, documentation, etc.) the original data contributors should be credited with authorship of the final data product. This provides an incentive for principal investigators to submit their data to the archive.
- *Background information.* Why was the data collected, who funded the project, and what connection do the data have to other (historic or contemporary) data collections?
- *Scope and purpose of the program.* These cover the specific objectives of the data-gathering effort, the statistical design of the sampling effort, and the hypothesis being tested, if any.
- *Data collection procedures.* This covers the detailed description of the data collection procedures, including analytical procedures, calibrations, processing algorithms, and so forth. Photographs of the collection apparatus and topographic maps of the collection area, for example, should be included in the documentation of the collection procedures when appropriate.
- *Station history.* If the data were taken at a station (e.g., meteorological data), then the complete station history for the length of the data stream should be provided.
- *Description of instrumentation.* This is the description of the instruments in data collection and processing, including make and model number (some investigators might require actual instrument serial number to be recorded).
- *Definitions of calibrations applied.* This describes exactly how the data were processed after collection, and will include standard processing procedures that may not be the norm 20 years hence.
- *Full variable definitions.* Define every variable in the data set and their connection to any other. Fully describe any special processing of variables and whether any are derived, averaged, or estimated values.
- *Definition of calculated variables.* Provide not only the description of how derived variables were calculated, but also formulas used and copies of the completed code used to implement the calculations (in case there were errors in the computer code).

- *Description of adjustments.* Describe any adjustments of the data set made by the data gatherer after the initial data collection. Data not in the data set may have to be described: for example, suspect values that were discarded before compilation of the final data set.
- *Quality assurance at the data center.* Describe the quality assurance procedures applied at the data center.
- *Modifications made at the data center.* Describe any modifications to the data set made at the data center, such as summarization or merging with other data sets.
- *Limitations of the data.* Every data set has limitations in its accuracy or appropriateness for certain uses. Explicitly acknowledge limits in precision, accuracy, and application, and work with principal investigators to identify and resolve such problems.
- *Data transport verification statistics.* Provide statistical summaries for every variable in the data set (e.g., number of observations, minimum, maximum values, average value) so that the user can ascertain whether the data were correctly transported to the user's system.
- *Full or sample listings.* Partial listings of voluminous data sets (a listing for every file) provide examples of how the data should look. Full listings of small data files provide data to those without electronic capabilities and provide an alternate method of archiving the data.
- *Input/output routines on the transport medium.* Provide listings and the actual code on the medium for the input/out routines to read the flat files into appropriate languages or applications packages (e.g., FORTRAN, SAS, ARC/INFO) in order to eliminate transcription errors.
- *References.* Provide references to papers describing the sampling process used to produce the data set or papers reporting on the analysis of the data set by other users. For gray literature, provide the full reprints themselves, because these may be hard to obtain in 20 years.

Level of Service	Level 1	Full set of data management functions in the GCDIS with packaged data set, complete with necessary ancillary data; documentation adequate for full use with confidence; priority given to reassembly, if necessary.
	Level 2	Full set of data management functions in the GCDIS; ancillary data identified; documentation is a detailed description with quality estimates; reassemble lower priority.

**Level 3      Minimal set of functions in the GCDIS; documentation is a summary of data sets.**

The IWGDMGC activities relative to the adoption of standards and guidelines for the documentation of global change data sets will be coordinated with other groups addressing similar issues, such as the FGDC.

**Reassembly**

The identification and priority-setting functions of GCDIS will produce an inventory of essential or high-priority data sets. It is expected, however, that repackaging or reassembling some of these data sets will be required to best meet the needs of the users. As one possibility, a data set could be reassembled by reformatting: that is, modifying its structure to conform to a more widely accepted standard. Also, the data set could be migrated to more efficient media that would permit wider distribution to the user community or merging with other data sets.

A specific example of reassembly services is represented by the need for the detection of significant long-term global change. As stated in a recent report by the National Research Council, "We believe that the detection of significant long-term global change is so central to the goals of the USGCRP, and so clear an obligation to future scientists, that it should be considered explicitly for added emphasis in the early stages of the program."<sup>2</sup>

This central requirement for the detection of global change, and the determination of its natural or human-induced origin, requires that long-term data bases of key parameters be established and updated regularly as the data and information from new observations become available. These special data bases each need to be of the highest quality and must combine the data and information from all the applicable observations into a standard, regularly produced, special data and information product with continuity from year to year. Examples include global ozone profiles, solar irradiances, cloud cover, snow cover, land cover, land and ice surface temperatures, and sea surface temperatures. All these special-standard products will have the functional service levels of essential priority data and information sets.

The importance of these special-standard products makes it particularly important that international agreements on both the individual parameters and on their formats and other aspects be reached as soon as possible. In addition to its need by the USGCRP, this special-standard product data and information activity will lay the

---

<sup>2</sup> National Research Council, Committee on Global Change, 1990. *The U.S. Global Change Research Program*, National Academy Press, Washington, D.C.

groundwork for the United States to help lead the development of such international working relationships and agreements, wherever appropriate.

## **Production of Derived, Value-Added Data Products and Services**

Data management by principal investigators and at data centers organized by scientific discipline are the traditional mechanisms used to maintain research data. These mechanisms developed to serve the needs of discipline-oriented researchers, especially in support of the geophysical sciences.

The advent of global change research has changed the way researchers investigate scientific issues. The issues of interest in global change research are some of the most complex ever contemplated. They go far beyond inquiries bounded by clear disciplinary lines. For example, assessing the effects of sea level rise on coastal regions requires data from more than a half-dozen disciplines to be integrated into one geographically registered data base. In response, agency data centers are adding new data and information products that are built with other GCDIS holdings.

In particular, existing issue-oriented Information Analysis Centers (IACs)—the Department of Energy’s (DOE’s) CDIAC, for example—provide these interdisciplinary services. The IACs proactively participate in the identification and creation of needed data sets from multiple extant sources to support specific issues. IACs support a user community beyond researchers, specifically including educators and the assessment staff that support policymakers.

The IACs focus on the integrated, value-added data and information products needed by both the research and policy-making communities to assess and mitigate specific global change issues. Whether global change IACs are formed as new entities within the GCDIS or incorporated into existing data centers, the establishment of IAC functionality is critical.

### **User Scenario—Model Development**

An example of the type of value-added, derived-data product that an IAC functionality would produce would be in support of researchers working with General Circulation Models (GCMs). GCMs are used to predict future global climate, and are based upon a number of assumptions. One of the parameters GCM researchers need is estimated future levels of atmospheric carbon dioxide (CO<sub>2</sub>). Among other things, future levels of CO<sub>2</sub> are calculated based on historic fossil fuel consumption and estimates of future fossil fuel consumption. The data set of historic emissions of CO<sub>2</sub> due to the fossil fuel consumption and other anthropogenic sources is a derived data set. It includes data on energy use from the United Nations Statistical Office, data on cement production (cement production involves substantial

amounts of CO<sub>2</sub> in developed countries) obtained from the U.S. Bureau of Mines, and data on well head flaring (burning of large amounts of natural gas at the well head—data not included in energy production figures) at oil fields from the DOE. These data sets must be in common formats, combined across time and country, and quality assured before fuel-specific algorithms can be applied that ultimately provide CO<sub>2</sub> emission estimates. These derived, value-added data sets are typically needed in the address of complex global change issues such as climate change, sea level rise, and socioeconomic questions. The data sets' production has been cited as a key component of the GCDIS in the NAS strategy report.<sup>3</sup>

### User Scenario—Policy Decisions

A further example of synthesis services relates to the immediate need for policy decisions on how to prevent or adapt to climate change. For these decisions to be based on the best available information, models that integrate smaller models of parts of the global change system are needed. There are many of these smaller models, including models of the world economy, atmospheric chemistry, ocean circulation, ecological systems, and social preferences and dynamics. The scale of some of these smaller models already strains the largest computers and makes it unlikely that it will ever be possible, or even desirable, to combine them all directly into an integrated assessment model. Rather, integrated policy assessments will need reduced-form data from these other models that captures their most salient features. Providing such reduced-form special data and information products is essential—not only for policy decisions, but also for a whole range of education and interdisciplinary research needs.

## Supporting Functions

### Browse Information

The browse function will provide low-resolution, summary, or visualization information about the data set. Since this information is discipline dependent, each implementation team will define the criteria and guidelines for data sets in its area of research.

---

<sup>3</sup> National Research Council, Committee on Geophysical Data, 1991. *Solving the Global Change Puzzle: A U.S. Strategy for Managing Data and Information*. National Academy Press, Washington, D.C.

## **Quality Indicators**

Quality assurance is of paramount importance in the creation of global change data sets. The data must be carefully reviewed for consistency and accuracy, with adjustments made for changes in operation or instruments when possible, especially in long-term monitoring data sets. This is important if we are to identify small-scale trends in long-term data sets. This quality assurance effort is shared by the principal investigators originating the data, the scientific projects under which the multidisciplinary data sets are being collected, and the data centers that archive and distribute the data.

Researchers must be the primary source of quality assurance for checking their data, since they are closest to the collection process and are most knowledgeable about problems with the measurements. However, researchers are often reluctant to quality assure and document their data to archive standards. This is understandable because they often do not need such high standards for their own use of the data, and especially because they are not rewarded to do so. (In some cases they may not be particularly good at it.) The project under which the data may be collected by many principal investigators has the obligation to enforce standards of formatting, definitions, documentation, and quality assurance among the various data streams being generated to produce a consistent, integrated data base useful in systems analysis. The archive data center also has a major obligation in ensuring the quality of the data being archived. It is the last chance to review data quality before distribution to the research community at large. It is at this step that shortfalls in the quality assurance process at the principal investigator and project level can be identified and corrected.

Experience has proved that few complete data sets produced by researchers are without problems. Often, quality assurance includes simple checks for missing or unreasonable values or inconsistent correlation. But a good data center will question further. Some large observational data sets are received unfiltered (for completeness); but by applying reasonable constraints on the quality of observations (always working with the principal investigator), they can be made beyond question.

The resulting data set may be smaller, but confidence in the reliability of the data is greatly increased. In the case of computer models, appropriate quality assurance measures include error and sensitivity analysis requiring special codes and research techniques. Of course, any changes made are in conjunction with and with full approval of the original researchers to ensure their appropriateness. The accompanying documentation clearly explains what was done during the quality assurance process.

Depending on the data involved, the effort can be extensive and often can be the single most expensive step in processing a data set for distribution. The cost of full-quality assurance for a large data set makes it imperative that priorities be set within each data center so that only the most valuable data sets receive the maximum level of quality assurance. However, for these pivotal data sets, this level of quality assurance is absolutely necessary since future research and policy decisions may rest on careful analysis of these important data sets.

Data set documentation should specifically identify what quality assurance procedures were applied to the data either by the principal investigator(s), the project team, or the archive data center. Data variables themselves may have quality tags associated with them to indicate the level of confidence placed in their value or to indicate known problems with measurements. The value and use of data quality tags will vary between data sets and the analyses to which they will be subject.

Level of Service	Level 1
	The data will be put through a system of real-world analyses similar to those they will be subjected to by the user community. Important data sets should be released as beta test versions to interested researchers who can exercise them thoroughly by applying actual analytical algorithms. Any problems can thus be identified and resolved at the data center (working in conjunction with the original suppliers of the data) before final release to the general user community.
	Level 2 Data centers will scan all data for proper ranges of values and provide simple logical testing—summing values that should have a known value (e.g., 100 percent), correlating related values (such as cloud cover and sunshine hours), or locating sampling points relative to known regions (oceanographic samples should be located in the ocean).
	Level 3 A data center will read all data sets for adherence to formatting conventions and proper identification of missing values.

## Outreach and Education

There are multiple aspects to the outreach and educational needs that the global change data and information management program must satisfy. Satisfying these needs will be one of the functions of the Content Subgroup. Examples of these outreach and education needs include the following:

- *Public information.* There is a broad spectrum of users who need global change data and information. Furthermore, the global change data and information themselves are in multiple disciplines. For the data and information to be used to the fullest, it is necessary that the total user community be made aware not only of the data and information themselves that are available, but also of the related services that can be provided. Special products will be produced to meet this need. The data and information management system will not be successful unless it is used, and it will not be used unless potential users know enough about it to use it.
- *K–12.* It is clear that learning about the global changes of our home planet is not only important, but also interesting and challenging at all levels of the educational system. Global change data and information help provide the fuel for such education. However, most of the data and information are not in forms that can be readily used at the K–12 levels or even by the general public. With the advice of K–12 educators, special attention will be given to the production of special data and information products suitable for use by K–12 teachers and students and by the media.
- *Training.* Clearly, the mandate of the global change data and information program to have data and information available in useful forms for the broad user community will require the efforts of an increased number of individuals cross-trained in both research and information science. Such individuals must also be professionally motivated to use their skills in the program. The data and information management program will encourage such cross-training through such mechanisms as graduate and undergraduate student programs at universities and visiting researcher programs at agency data centers. Such education initiatives also will help develop the professional acceptability needed to draw other talented individuals into both providing their data and information and to putting it into useful forms.

## National Policy Issue Data and Information

The Content Subgroup will ensure that the necessary data and information are available to resolve the important global-change-related policy issues the United States faces. Two examples of such policy related issues are shown in Figure 3. Other policy-related issues are being identified by the CGED.

As a specific example of GCDIS content system implementation, the Content Subgroup will assume responsibility for the coordination necessary to identify a limited number—a dozen, more or less—of such policy-related issues. This will require coordinating with multiple groups, including other elements of the USGCRP and the NAS. For each such policy issue established, the Content Subgroup will form

an ad hoc panel that includes representatives of the appropriate NAS groups, agency program managers, and agency data centers. It will be the responsibility of this panel of experts to identify

- The priority data and information that are potentially available,
- The data and information for which the existing level of service is below the requirement,
- Holes in the data and information coverage that need to be filled both with existing data and information and with their long-term continuation,
- Special products required, and
- The data and information management actions and schedules necessary to meet the imperatives of a particular policy-related issue.

This approach—working group interactions among advisory-, management-, and working-level experts that span particular issues—will help ensure that the best data and information will be available so that progress can be made on problems of central national importance. Further assurance will be provided by giving the priority policy-related data and information identified by this approach the highest level of service, level 1.

A pilot study has been initiated to test and demonstrate the applicability of this approach for GCDIS content coordination on an actual policy issue before implementing this approach on the full range of USGCRP policy issues. Because of the imperative to fill the planning gap on requirements for CO<sub>2</sub> emissions monitoring that is an integral part of the Climate Framework Convention, the issue selected for the pilot study is the ability to determine regional sources and sinks for atmospheric CO<sub>2</sub> for application as a part of a monitoring system for a GHG emissions reduction agreement (see Figure 3). Based on preliminary estimates of the data and information needed to investigate this problem that are presently available through processes such as those outlined in Figure 3, it seems likely that one immediate result of the pilot study will be the identification of the additional data and information management efforts that must be mounted for the determination of the regional sources and sinks of atmospheric CO<sub>2</sub>.

## **Emerging Trends**

In the course of GCDIS implementation, the richness of approaches to the analysis and interpretation of global change data and information will continue to increase. Currently, three critical trends are emerging: the increasing use of model output from one scientific domain as input into the study of another domain, the use of data to drive data assimilation models in order to interpolate among measurements that are

relatively sparse in space and time, and the increasing dependence of field campaigns on the data from existing ground-based and satellite assets that are managed by a variety of agencies. These activities and others that can be expected to arise as the study of global change matures will cause major shifts in the way in that global change data are approached and the assets needed to support their effective utilization. The subgroup will work with other elements of the USGCRP to address these issues.

These three trends lead to the need for early consideration of several issues. The ways in which model output is captured and preserved will be critical. Learning from the Program for Climate Modeling Diagnosis and Intercomparison, the GCDIS will address guidelines for what data and information need to be preserved from particular models, as well as which models will have their output regularly preserved. Data assimilation is becoming an analytic method of choice. Not only will it be necessary to address the preservation of assimilated data products, but the extensive computational assets needed for a global data assimilation will increasingly become a necessary aspect of the data access question. Specifically, without assimilation, many data will be far less valuable.

Finally, the data assets of the various national and international agencies have value beyond their support of researchers seeking access in the sense of a library or archive. In the coming years significant field programs, such as the World Ocean Circulation Experiment (WOCE) and GEWEX, will mount major field campaigns that will depend heavily on these data assets. They will depend on other data not only for the postcampaign analysis, but also in support of the field operations during the experiments themselves.

## 5. Access System Implementation by Functional Areas

This chapter discusses implementation of the GCDIS access services, including an overview of the GCDIS access system, a discussion of how GCDIS access function implementation will be coordinated, a summary of the GCDIS access services, and discussions of each GCDIS access system functional area. This will be followed by discussions of the relation between GCDIS access functions and State, local, and international programs with which GCDIS must work.

For the GCDIS, the definition of system embraces the full range of people, procedures, hardware, software, and infrastructure (e.g., networks, telephones, mail) that together constitute a system providing users with access to data and information services (a mix of online and offline).

### Access System Overview

The GCDIS is the set of individual agency data and information systems: a mix of manual and automated, online and offline functions, supplemented by a minimum amount of new crosscutting infrastructure, and made interoperable by use of standards, common approaches, sharing of technology, and coordination of data policies. Thus, each agency's own data and information system will be that agency's component of the GCDIS.

The GCDIS comprises individual agency systems that must evolve to meet changing agency user needs and agency missions, and to take advantage of emerging information system technology. Such evolution will pose a major challenge to the agencies to ensure that it does not disrupt the interoperability that allows the GCDIS to exist. Also, changing needs of global change users and technology will create an evolutionary pressure to which agencies must respond while protecting agency missions and users. The coordinating mechanisms and the agency implementation approaches, including the standards used, must allow flexibility and change.

### Conceptual View of the GCDIS Access System

The conceptual view of the GCDIS access system is in marked contrast to the present system where, for the most part, individual agencies service only their own user communities. Although with the GCDIS each agency will continue to serve its own user community, agency data and information systems will be interconnected.

Users accessing any of the agency data and information systems making up the GCDIS access system will be able to obtain a comprehensive view of common GCDIS access system services (common across the GCDIS agencies), while also continuing to have access to the specific services of that agency (which are tailored for the user community specific to that agency according to the agency's own individual mission requirements). The actual physical configuration within this conceptual framework will evolve dramatically as user needs and technology change and as the different agencies attain increasing degrees of GCDIS interoperability.

The system will also evolve in terms of the number of participating agencies. The IWGDMGC will actively work to broaden GCDIS participation to include the other agencies with data and information needed for the USGCRP. To encourage their participation, the GCDIS is planned as a system that an agency can join with only a modest commitment and then work toward a fuller implementation of the information service objectives of GCDIS. Although the desirable level of enhanced participation will involve provision by the agency of the full suite of metadata appropriate for their data and information, the minimal requirement for agency participation will be

- Availability of data or information relevant to global change,
- Directory level metadata describing these data or information,
- A credibly permanent commitment to maintaining these data and information and making them available,
- Interoperability with the rest of GCDIS sufficient to make their directory level information accessible from any other GCDIS entity, and
- User support for the data or information by somebody.

There is now and always will be significant diversity among and within agencies regarding their data and information systems, ranging from highly automated systems to largely manual systems. The evolution of systems within agencies will proceed at different rates, given agency user requirements, mission needs, and budget constraints (this plan assumes that the GCDIS access system implementation must be accomplished within currently projected agency budgets). The GCDIS access system approach must be sufficiently flexible and adaptable to function in such a highly heterogeneous and changing environment.

### **Interoperability of Services**

From the users' point of view, the interoperability required for the GCDIS access system implementation to be successful must result from a sufficient degree of commonality among agency data and information systems and services.

Commonality does not imply major changes to the data and information systems. Locator technology freely available on the international Internet will be used to provide a common look and feel at a general level for systems that differ radically from each other, although the underlying systems may be quite different in supporting more specific functions. This enables users to view the overall aggregate global change data base and supporting services as a coordinated whole and yet still have the option of accessing agencies' data and information services by the methods defined by each agency. Sufficient commonality in this context can not be defined *a priori*—continuing involvement of global change users in advising, testing, and evaluating the evolving implementation of the GCDIS access system will be essential to ensure that the degree of interoperability attained by the GCDIS access system meets their needs. The agency elements that will become agency components of the GCDIS access system are and will be a heterogeneous collection, ranging from almost entirely manual systems to highly but never completely automated systems. Interoperability must be addressed in terms of common standards and guidelines for manual procedures as well as information system functions.

There must be interagency *coordination* on the common standards and approaches required to meet the needs of the global change user community, while also meeting the needs of agency-specific user communities. Sufficiency will be measured by users evaluating the usefulness of the GCDIS access system.

Underlying the interoperability seen by the user will be a variety of implementation approaches, such as use of a client–server architecture that will evolve with advances in information systems technology. The pace and extent to which individual agencies are able to implement interoperability will vary with the agency's own mission, schedule, and cost constraints. Attainment of the goal of making disparate services and systems look like a common logical system to the user will, therefore, not be achieved uniformly across all agencies. Thus interoperability and the GCDIS access system gradually will evolve.

### **Standards and Procedures Approach**

The GCDIS will adopt recommended common standards and procedures in areas critical to the consistency and coherence of the global change data and information holdings and supporting access services. The purpose of reliance on standards is not only to facilitate use of the system and its products by users but also to facilitate evolution of elements of the system by building in technology independence through adoption of such standards as the Portable Operating System Interface (POSIX), the Spatial Data Transfer Standard (SDTS), and related approaches.

The Standards Subgroup will provide support to all the IWGDMGC subgroups. Examples of support will be to provide information on training, documentation, and

contacts; to be the focal point for the coordination of recommendations on standards for agency implementation plans; and to promote the development of new standards where needed. The Standards Subgroup will submit these recommended standards to the IWGDMGC for adoption.

### **Standards**

In the functional areas described in this plan, the objective will be to identify and use existing standards, whenever possible, rather than to develop new standards. Where FIPS either do not exist or are not applicable, preference will be given to adopting international and national standards developed by recognized voluntary standards bodies such as the International Standards Organization (ISO) and the American National Standards Institute. If no FIPS or voluntary standards exist for an important area of need, and if timely development by existing standards bodies seems unlikely, then consideration will be given to establishing an interagency standard to meet the need.

### **International Standards**

Implementation of the GCDIS access system by U.S. agencies will facilitate international cooperation by making use of international standards and protocols (e.g., the Unicode and the ISO standards, the CEOS standards) wherever practical. The U.S. GCDIS access system will be a node on an international global change data and information system.

### **Procedures**

Procedural guidelines in areas such as user services and billing and other related administrative functions will be adopted and implemented to improve services important to the global change researcher.

### **Technology Sharing among Agencies**

Technology sharing initially will focus on software needed to provide commonality required to reduce differences among agency data and information systems. The exchange of software will simplify interagency mediation needed to bring about the interoperability needed for the GCDIS access system. Technology sharing also may offer cost savings for agencies and facilitate more rapid implementation of needed capabilities. Bilateral or multilateral technology sharing will expedite implementation of the GCDIS access system.

Software will be shared among agencies participating in the GCDIS access system. The software will be fully documented and will embody architecture and design approaches, such as layering and use of standard interfaces to avoid hardware

dependence or vendor dependence, and to allow modification or addition of modules without redoing architecture. The agency accepting software will be responsible for implementing the software in its own environment, establishing the capability to support it, and adapting or extending the software as needed to meet the agency's own needs while maintaining the agreed degree of commonality needed to support the GCDIS access system. The agency providing the software will provide full documentation, technical assistance, and support to the extent possible within its program constraints, but the receiving agencies will need to have or develop their own capabilities.

## **Coordinating Access Implementation**

### **Access Subgroup**

Data access issues will be coordinated by the combined efforts of the Catalog, Networks, Standards, and Libraries Subgroups. The Catalog Subgroup will act as the IWGDMGC overseer of the GCMD. The Network Subgroup will analyze requirements for internal and external networking required for the success of the GCDIS access system.

The Standards Subgroup will develop recommendations for FIPS and voluntary standards to be adopted in the functional areas described in this plan. These recommendations will be submitted to the IWGDMGC for adoption. Agencies will decide whether to adopt the recommendations and the appropriate time frame for their implementation. An overall decision to accept a recommended standard will be by agreement of the agencies involved in the GCDIS access system implementation.

The IWGDMGC Library Information Subgroup will coordinate the identification, organization, creation, and distribution of library and information products and documents related to global change research, referral, and educational activities, and serve as a clearinghouse that coordinates library and information center collection efforts among organizations involved in global change research.

Other GCDIS access system functions will be implemented entirely within the participating agency elements, with specific coordination of existing implementation efforts to be accomplished through the mechanisms discussed in the next section. In addition, the Access Subgroup will include a catalog and distribution services team, which will include representatives from agency implementation efforts, to ensure that these efforts are coordinated across the agencies.

### **Agency Implementation Coordination**

Agencies in the process of upgrading or developing their agency data and information systems as their components of the GCDIS access system will coordinate the technical details of the design and implementation of their systems within the framework provided by this implementation plan. Agency representatives participating at this level will be directly involved in each agency's data system implementation effort; these people are in the best position to arrive at a consensus on specific technical recommendations to the GCDIS access system agencies. They also can help establish appropriate links to other interagency data and information access programs.

Each agency will establish its own mechanism within its own implementation effort, inviting participation of agencies engaged in parallel efforts, to make recommendations concerning specific design choices or implementation decisions to allow the interoperability required for the GCDIS access system. Each agency should then incorporate the GCDIS access system recommendations into its own systems to the extent possible, given cost and schedule constraints.

### **Access Services**

This section defines the system functionality required for GCDIS access services. Access services include a complete range of data and information services to users—catalog, order, browse, distribution, and user support—and underlying archive and network services. Electronic access to these services will be provided to users via the Internet. For the GCDIS access system to succeed, participating agencies must cooperate to achieve interoperability among their various methods of providing these services. That interoperability will allow users to see a consistent set of services spanning the agencies; users will be able to identify, order, and receive ensembles of data and information products held by some or all of the agencies. Their use of these products will be facilitated by common media and formats and consistent documentation.

Some agencies' participation in the GCDIS access functions will be limited. In particular, DOD participation in the GCDIS is limited to making unclassified and unrestricted data and information available to appropriate civil agencies; GCDIS access functions for DOD data and information will be assumed by those other agencies.

Although it is a driving assumption that development and extension of electronic, online capabilities is needed for GCDIS access services, GCDIS access services always will be provided by a mix of manual and automated, online, and offline

functions. Whereas that mix might include more electronic services in the future, the need for manual services in such key areas as user support will remain. Thus, the broad definition of *system* in the GCDIS is a combination of hardware, software, people, and procedures.

A user view of the functional areas and an overall schedule establishing goals for progressively achieving interoperability in these areas will be presented. This will be followed by a discussion of each functional area in detail, including definitions of levels of implementation of the functionality within each functional area, and milestones for progressive, phased implementation by the agencies.

Therefore, the overall plan for the GCDIS access system implementation will be presented as a combination of efforts for the individual functional areas, and within each area implementation will be addressed on an agency-by-agency basis, with individual agency implementation plans set in the context of the overall schedule and priority.

### User View of the Functional Areas

The following is a presentation of the GCDIS access system functional areas from a user point of view. This is not intended to suggest a strictly sequential approach to implementing the GCDIS, but to provide guidance based on user needs for phased and overlapping progressive implementation of the GCDIS access system. There is a fundamental requirement for user support services, provided by expert personnel, accompanying any and all access functions, regardless of the degree of automation of the functions.

- *Archive*. The user must be able to rely on the long-term integrity of priority global change data and information, and required documentation must be preserved.
- *Offline access*. At the most basic level, the user must have access to priority global change data and information, including full hard-copy documentation of all data sets and related reports and publications, in an offline, manual mode.
- *Directory*. At a minimum, the user must be able to identify readily and locate priority global change data sets, data systems, and related information across GCDIS agency global change holdings and related holdings of external organizations (national and international).
- *Inventory*. The user must be able to identify and select, using (at a minimum) common spatial and temporal references and the particular combinations of data and information from multiple agencies needed for the research effort, drawing upon priority global change data sets held by the agencies.

- *Order.* Having identified needed data and information, the user must be able to order it.
- *Distribution.* Timely delivery and common formats offer important advantages to users by reducing the time between when a data set (and especially combinations of data sets) is requested and when the user can begin actual research with the data set(s).
- *Browse.* This allows the user to screen data sets to minimize the amount of data actually ordered, thus saving the user time in sifting through whole data sets instead of browse samples or subsets, conserving GCDIS capacity.
- *Guide.* Online as well as hard-copy availability of detailed documentation ancillary to key data sets provide essential information in convenient form.
- *Information.* This means access to and integration of library and information center collections, services, and information products (such as a wide array of published research in the form of books, journals, conference proceedings, technical reports, maps, geographic information systems, and multimedia, as well as gray literature, metadata, and small data sets). Identifying and creating access to papers relevant to particular global change data sets will receive special emphasis.
- *Billing and administrative.* Possible implementation efforts in this area will be considered in the future.

Networking is an underlying requirement for all these functions, except offline access. Each functional area requires network support, therefore network requirements will be addressed for each functional area.

International access is the extension of these services to provide access to global change data sets held internationally and, in turn, to allow international access to GCDIS functionality. The program plan calls for implementation of international access during the period 2001–2010. Some GCDIS access system agencies are now actively involved in international efforts that will be broadened in the future.

### **Access Function Implementation Link to Data Set Priorities**

Resources for implementing GCDIS access functions will be limited. It is essential that data sets of greatest importance to the USGCRP receive the highest levels of support. Table 4 maps the level of implementation for each access functional area against the categories of data set priority, and is a companion to the similar table in Chapter 4 that maps content areas against data set priority. The table should be interpreted as establishing a goal for a floor level of implementation in each case. For example, an inventory implementation level of 2 for a high-priority data set would be

the minimal goal for the GCDIS access system. The access functional areas and the levels of implementation for each will be discussed in detail in later sections.

In the case of the directory function, the levels of implementation are not readily correlated with data set priority. Rather, it is most urgent that the directory information be assembled and made available for the most important data sets; priorities for these efforts will be consistent with data set priorities.

**Table 4. Levels of Implementation  
(GCDIS Access Functions)**

**Functional Areas**

Data Set Priority	Archive	Guide	Inventory	Browse	Order	Distribution
1	NARA/NIST off-site backup	Online, interoperable	Online, interoperable	Online, dynamic & static	Online, GCDIS-integrated	Standard media and formats
2	NARA/NIST on-site backup	Online	Online, GCDIS standard	Online, static	Online, manual forwarding	Add electronic delivery
3	NARA/NIST compliance	Offline, GCDIS standard	Hard copy or manual services, GCDIS standard	Offline, digital	Manual forwarding	Standard media
Not Rated	Controlled storage	Existing hard copy	Existing hard copy or manual services	Offline, hard copy	Manual forwarding	Standard media

### **Access Function Implementation Link to Type of User**

The wide spectrum of the user community includes users with

- Access capabilities ranging from mail to telephone to the Internet,
- The ability to handle data and information from paper only to optical disk and electronic, and
- Very limited technical knowledge of a particular discipline to expertise in that discipline.

The GCDIS will recognize and meet these disparate user access requirements. The technical capability for meeting such a spectrum of needs in the first two areas is well developed and available. To meet the needs in the last area, it will be necessary to build special directory capabilities to help guide specific classes of users to the data and information products that they require.

For example, educators in the K–12 levels will need to be directed to those products suitable for their use without being immediately confronted upon entry into the GCDIS with the total contents of the GCDIS. Similarly, many of the users interested in policy aspects will need to have identified data and information products that can either meet their needs or help guide them to more detailed information.

A final example of the need for these special directory capabilities is that of researchers in one discipline wanting to work on interdisciplinary problems. Many such users will need special help in finding summary-level information in disciplines outside their own to guide them in their more detailed use of the GCDIS.

## User Support and Outreach

### Description

User support services include providing expert staff personnel to assist users with any of the GCDIS access system functions. Regardless of the degree to which the GCDIS access system functions become automated or accessible online, a requirement will exist for user support personnel to interact directly with users, answering questions, and providing advice and guidance. This function also includes preparation and distribution of user aids of various types.

The data or information center should reach out to the researchers, policy makers, educators, and the general interested public that make up the user community. Data and information product announcements in scientific journals, flyers to discipline departments at universities, and maintenance of mailing lists that include attendance lists from major workshops or meetings are effective ways of informing the user community of a center's holdings and services. A periodical center newsletter is one means to establish a stable audience for center communication.

Just as it is important to reach out to users to inform them of the system capabilities, user feedback is critical to the growth of the global change data and information program. The components of the system of user services, directory or catalog offerings, networks, and procedures must be capable of responding to such changes as increased user sophistication, priorities' changes, and technology advances. The direction of these changes often come from the users. Information about problems experienced by users can provide valuable guidance on system technical

shortcomings. Feedback through user forums and questions that come into information centers offer indicators about needed data and information lacking in the system. These and other techniques, such as automatic collection of access system search statistics, will be implemented to guide enhancements to the overall system.

### **Needed Interoperability**

The user support staffs of the various agency GCDIS access system elements will exchange information about their organization's data and services so that a user contacting the user support staff of any GCDIS element can obtain information about all other GCDIS elements, and obtain referrals to GCDIS access system elements offering data and services of interest to the user.

### **Agency Applicability**

This function applies to all agencies (or their designees) participating in the GCDIS access system at any level.

### **Agency Implementation Overview**

Continual coordination of agency user support functions will be provided through the standing User Support Team, composed of agency user support staff with active participation of members of the User Advisory Group and other user representatives.

## **Networks**

### **Description**

Each GCDIS agency will establish a suitable level of connectivity to the Internet, a global network of networks connecting the research and education communities throughout the United States and to nearly 60 (as of March 1993) other countries. In the United States, the Internet consists of NSFNET, regional and State networks, agency networks, and local networks such as campus networks. The future development of the U.S. infrastructure will be strongly influenced by the NREN program, the networking component of the HPCC program, and the NII program. Connections to this networking infrastructure are required to support external user access to GCDIS access system services and internal operations between the GCDIS access system components needed to support user services.

This set of connections is referred to as the GCDIS access system virtual network. Connections on the GCDIS access system virtual network will be set up to handle the types of operations that may take place to support GCDIS functions accessed by a user at one GCDIS access system node. For example, if a user connects to agency A's system and performs an inventory search spanning the inventories of agencies B and C as well as A, some communication between agencies A, B, and C will be

necessary to relay the search request to the inventories and then transfer the search results back to agency A for assembly and presentation to the user. The nature of the communications between the agency systems will be a high frequency of small transactions that require very rapid response in order that the end-to-end or net response time to the user be rapid enough to meet user needs. Also, in some cases, the connections will be needed to support other global-change-related functions, such as supercomputer access.

In the case of both external and internal network requirements, an analysis of the frequency and size of the transactions that the network must support will be made. Response times will be established that meet user needs. Attention will be paid to peak as well as average capacity requirements (e.g., in terms of concurrent user support).

### **Needed Interoperability**

In order to support open electronic access by users or electronic exchanges with other GCDIS access system elements (in the same or other agencies), agency GCDIS access system elements must have the appropriate network connections, including Internet access. The capacity of the agency elements network interface must be sized to support the GCDIS access system functions requiring network support.

### **Agency Applicability**

The need for networking support as an integral part of the GCDIS access system applies to all agencies (or their designees).

### **Agency Implementation Overview**

Agency plans for implementing online GCDIS access system functions will include implementation of required network support. For example, agency plans for implementing online GCDIS access system inventory services will include identifying agency elements involved and establishing network connectivity for those agency elements, and sizing the networking capacity to meet the needs of the inventory function.

Three levels of network user performance requirements have been identified that are needed to provide access to GCDIS user services:

- *Network user performance level 1: Online data services.* Network user performance level 1 supports online dynamic science, which means transfer of data in real time, access to supercomputer computational models, and scientific conferences. This level of network performance requires sustained transfers of up to 1 megabit per second (Mbps) or more. This network performance level is provided with either dedicated Ethernet and T1 services

(1.544 Mbps and higher), or shared fiber-optic networks (e.g., a fiber-optic digital data interface local area network [FDDI LAN] with T3 or switched multimegabit data service access, or asynchronous transfer mode).

- *Network user performance level 2: Online browse.* Network user performance level 2 supports user access to precomputed browse products that are stored online or near line. Near-line data are staged dynamically into the online hierarchy at the time of an initial preview request, so that whereas the initial browse product may not be transmitted back to the user for as long as a minute, subsequent products in the same data set may be sent within a few seconds. This level of network performance requires sufficient bandwidth to support burst transfers of up to 100 kilobytes per second (Kbps) for up to 1 or 2 seconds. This network performance level is typically provided with shared Ethernet connections using T1 access lines.
- *Network user performance level 3: Metadata access.* Network user performance level 3 supports user access for directory, guide, and inventory searches of metadata. The metadata typically are text records accessed with brief text queries. This level of network performance requires sufficient bandwidth to support a number of short interactive messages with rapid response times. This network performance level typically is provided with voice-grade telephone line access (i.e., direct modem connection at 2.4 to 14.4 Kbps, or shared Ethernet connection via 56-Kbps access line).

In addition, networking between GCDIS access system elements is also required to support GCDIS user services. Within a given agency, its GCDIS element might be a collection of agency elements linked together as a single but distributed agency GCDIS element, or a set of agency elements, each functioning as separate GCDIS access system nodes. Internal GCDIS access system networking refers to networking between the GCDIS access system nodes required to support GCDIS access system user services. These internal GCDIS access system network services are required to support the rapid searching and relaying of metadata and data in support of services provided at each agency's GCDIS access system node user interface. The nature of network communications within agencies and among agencies in support of each of the network user performance levels is similar to the levels seen by the users, except that in some cases there may be more system-to-system interaction than seen by the user, and that the internal GCDIS access system networking must be able to support a higher level of concurrent use than individual GCDIS access system node concurrent user support (i.e., concurrent users at a number of GCDIS access system nodes might invoke GCDIS access such system services as inventory searches that each spread out across all GCDIS nodes).

## Directory

### Description

The GCDIS directory will provide information to its users on all priority, global-change-related data and information that are available and that have sources identified for user help with both acquisition and use. This applies whether the source of the data and information is international, National (agency), State, local, or from an individual.

The GCDIS directory consists of several elements. At the core is the GCMD and associated Directory Interchange Format (DIF)-based directories, such as the Consortium for International Earth Science Information Network (CIESIN) directory, which contain descriptions of data sets relevant to global change. The descriptions are provided by the agencies in a standard DIF which is automatically loadable into the DIF-based directory data bases. Conformance to DIF allows this core information to be searched by standard methods and to return data descriptions that are consistent in the type of information content. The DIF-based GCDIS directory function makes possible coherent searches across agencies by providing functional consistency in terms of space and time references and terminology used to name and describe data.

The DIF is one of many formats that are used to convey locator information, and DIF can be converted to and from the Government Information Locator Service (GILS) core elements format without loss of information. This allows DIF-based directories to be easily incorporated into many other directory services, and also allows DIF-based directories to bring in entities from other directories throughout the world. To accommodate existing library information systems, the DIF also supports a representation in the U.S. Machine Readable Cataloging (USMARC) format. Transformation from the DIF format to the USMARC format, however, is not easily reversible because the DIF format has more specificity than is currently defined for the USMARC format.

A number of data set directories for agency data and information existed before the creation of the DIF-based directories. These are the second element of the overall GCDIS directory. None of these directories are specifically oriented to the topic of global change, but many have relevant data descriptions. The information content varies in format, so it is not always easy to obtain comparative information. Nonetheless, the directories do represent a valuable resource and need to be made searchable through the GCDIS. These directories are searchable by the standard methods used in the past, but it is important also to be able to search multiple directories simultaneously for relevant data and information. The present way to do this is to use network tools such as Gopher and Wide-Area Information Server (WAIS) to allow access to these directories and make their information content

searchable by standard text retrieval and field-searching methods. The DIF-based directories can also be made searchable by these methods, but these methods do not presently allow the user to know valid values that can be entered in some of the controlled fields. Consequently, it is important to continue the work of describing high-priority global change data sets in the DIF-based directories.

Information other than data set descriptions, such as reference documents, regulations, study papers, and so forth, also needs to be locatable through the GCDIS directory. This could include, as well, pictorial materials such as maps and sample images. For this much broader amount of information, the GCDIS needs to function like a digital library and use the techniques that have been and are being developed for this purpose. Abstracts or summary descriptions for these materials (or even the documents themselves) can be retrieved through the GCDIS directory using the network tools mentioned earlier. The GCDIS directory will also lead the user to the standard online library search capabilities wherever available to the user community. The data set directories can be considered as one small part of this very large body of information.

Finally, the GCDIS directory must provide pointers at the guide and inventory level to other data and information systems and allow the user to gain more detailed information about global change data or even the data themselves. In many cases the GCDIS directory provides automated connections to these other systems. The user must know or be told, however, that these systems can be accessed directly and that it is not always necessary to go through the GCDIS directory to use them. In the same way, the other systems can make known to their users that the GCDIS directory exists and, perhaps, offer an automated connection to it to get directory-level information about data being pursued.

The GCDIS directory will be accessible by networks and dial-in lines from around the world. The directory will contain information from around the world and will not be limited to just U.S. entries. There will be at least one distributable version of the DIF-based directories available for stand-alone use on personal computers that retains as much of the functionality of the online versions as possible.

The DIF-based directories are part of the CEOS International Directory Network (IDN), which is a federation of directory systems from around the world bound by sharing information about data in the Earth and space sciences through the exchange of DIF files. The GCMD is the U.S. coordinating node of the CEOS IDN. There are two other coordinating nodes in Italy and Japan that are the European and Asian coordinating nodes, respectively. These coordinating nodes are gathering points for information exchange within the CEOS IDN, and each of them have complete and identical data bases and software that are kept consistent with each other through

biweekly automated information exchanges. Thus, the GCMD receives global change data information from international sources as well.

The GCMD and the CIESIN directory employ client-server architecture, with multiple clients accessing distributed servers. The DIF-based directories may evolve to become a single virtual GCMD directory composed of a federation of DIF-based servers, perhaps one for each participating agency, which can all be accessed through queries initiated by client software. These would support standard national and international protocols for information search and retrieval, such as ANSI Z39.50 and ISO 10163. This possible evolution of the DIF-based GCDIS directory function is now being tested with the GCMD and CIESIN servers.

### **Needed Interoperability**

Agencies describe their priority global change data and information holdings according to GCDIS directory standards for scope, content, space-time references, and terminology used in names and descriptions or abstracts. The directory function should be interoperable with guide and inventory functions as described previously.

Four levels of contribution by agencies to the GCDIS directory are defined as follows:

- Level 1. Agencies will operate a directory node containing information prepared to GCDIS directory standards describing agency global change data and information. This will include the resources necessary to keep the server's information up to date about the data and information resources to which it points. If a distributed server approach is adopted for the GCDIS directory, each agency directory would be one of the servers in the distributed GCDIS access system directory. If a centralized approach is continued, each agency node will contribute information to the central, coordinating node. This node will hold all information contributed by all of the agencies. This also will include the connection of underlying information systems to the distributed GCDIS directory using interoperable techniques to the extent possible.
- Level 2. Agencies contribute directory information for all global change data and information, prepared to GCDIS directory standards, to be added to the GCDIS directory data base.
- Level 3. Agencies contribute directory information for priority global change data and information, prepared to GCDIS directory standards, to be added to the GCDIS directory data base.
- Level 4. Agencies provide existing documentation of priority global change data and information to the GCDIS directory staff personnel, who in

turn prepare directory entries and add these to the GCDIS directory data base.

### **Agency Applicability**

The GCDIS directory applies to all U.S. agencies holding global change data and related information or supporting global change research and related data generation in outside agencies. The U.S. agencies work with their international counterparts to ensure that relevant international information is included in the directory function as well.

### **Agency Implementation Overview**

The present directory implementation activities for the GCDIS access system are as follows:

- NASA will continue to operate and develop the existing GCMD client and server;
- NASA will continue to develop and distribute the PC-based version of the GCMD;
- NASA will seek and review agency, academic, and other source global change data descriptions for inclusion in the GCMD; and
- NOAA will operate a Gopher server providing a structural hierarchy, access paths, information, and automated connections to the GCDIS directory and other parts of the GCDIS.

Agencies will continue to provide global change data set information descriptions to the GCMD by adding to and updating their data set descriptions in the present GCMD server. Useful and usable data sets will be described in DIF files and the DIF files loaded into the GCMD (in either the NASA or the CIESIN server). Adjunct products to the online GCMD, such as printed versions of selected information and CD-ROMs with information, data, and incorporated browse capability, will continue to be developed and distributed as resources permit. If the GCMD evolves to include multiple servers, agencies also may contribute information conforming to GCMD standards for new servers or provide access to servers containing agency directory information conforming to GCMD standards. All agencies will identify any data and information systems that can usefully be accessed through the automated connections (links) from the GCDIS directory level. Developments necessary to make these systems more interoperable will be carried out as resources permit.

## Guide

### Description

The guide function of the GCDIS access system provides detailed information about all agency global change data sets, to allow users to assess suitability of data sets for their research or application or to aid users in the use of the data set. This information includes bibliographic references to publications relevant to the data set and, in some cases, the abstracts or actual publications.

The guide function of the GCDIS access system will make possible coherent searches across agencies by providing functional consistency through including a standard set of topics (as they apply to each data set) and consistent terminology (e.g., data set references). There will be agency interfaces to agency-based guide systems that may include other agency guide information, or access or point to other agency guide systems or information.

Guides accomplish several functions within the GCDIS access system. Guides augment the general information in the directories with specific information about the data, such as detailed descriptions of the instruments and platforms used to acquire the data. This information permits the user to determine the availability and appropriateness of the data. Any information critical to the use and understanding of the data is collected in the guide. Explanations of projects under which the data were gathered also should be included. Short-term instrument outages, specifics about the recording instruments, and other factors that determine the suitability of the data are included.

The guide also fulfills the need to document carefully information about how the data set was derived. Detailed instrument logs, including instrument location, response, calibration and external factors affecting the instrument, may be included.

Information regarding the platform, such as description of satellites, stations, and observatories, also may be documented. Discussion of algorithms used to process the data must be detailed. Any information essential to the complete understanding of the data must be documented in the guide. This type of information is used to validate the data, and thereby the quality of research results. Without this information the data are useless.

In the case of data sets consisting of multiple granules (see the inventory functionality—page 66), the guide will include a generic description of a granule for that data set (i.e., the content and documentation included with each granule). The guide will describe any subset creation or other options available to a user for the given data set.

### Needed Interoperability

Functional consistency for the GCDIS access system guide requires a degree of consistency in the guide information across all data sets. This is limited to a standard set of topics, which will be addressed for each data set where they apply, and to standard terminology for terms such as data set names and parameter names.

The guide function will be interoperable with directory and inventory functions: that is, to allow a user to access guide information corresponding to a given directory entry, or to allow a user to access guide information concerning a data set that contains granules (page 52) meeting the user's inventory search criteria.

Four levels of agency contribution to the GCDIS access system guide function are defined as follows:

- Level 1. Agencies provide access to their online guide information for guide searches originating from other agencies' online guide systems, and support online access to other agencies' guide systems for their own guide users. This requires network connectivity for guide sites and network performance level 1 (57.5.5) for user access and internal GCDIS access system communications supporting the guide function.
- Level 2. Agencies make available online GCDIS access system standard guide information. This requires network connectivity for guide sites and network performance level 3 (page 59).
- Level 3. Agencies make available in hard-copy form standard guide information about the GCDIS access system (i.e., addressing GCDIS access system guide topics as they apply, and using GCDIS access system standard terminology).
- Level 4. Agencies make available in hard-copy form (e.g., user guides or handbooks) existing detailed documentation of global change data sets.

### Agency Applicability

The guide function applies to all agencies that hold global change data sets, unless for a given data set the directory level description constitutes a full description meeting the intent of the GCDIS access system guide functionality.

### Agency Implementation Overview

Each agency will independently implement its GCDIS access system guide functionality according to its own approaches and methods, including establishing network connectivity at guide sites where needed. This implementation overview will indicate when the various levels of GCDIS access system guide functionality will be

available at a user or interagency interface for each agency element separately implementing that functionality.

## **Inventory**

### **Description**

The GCDIS access system inventory function provides the user with the ability to identify particular combinations of parts of data sets (or whole data sets if the data set is most useful as a whole), ranging from one to all agencies, that meet user-provided search criteria (e.g., common space-time reference, or selected descriptors, information internal to the content of a data set that either distinguishes between parts of a data set or otherwise is important to the use of a data set).

Where a directory entry provides a summary description of an entire data set, and a data set guide provides detailed descriptive information about an entire data set, a data set inventory provides detailed information about individual parts of a data set in those cases where single or multiple parts of a data set are of greatest use to users rather than the entire data set. The individual parts of a data set are called granules. The decision about what constitutes a granule for a given data set is based upon patterns of use of the data set. The goal is to provide a user with the most meaningful way to identify and ascertain the availability of specific data to work with. A granule is a logical subset of a data set (or a whole data set if the data set is most often used as a whole) that may correspond to a physical subset of a data set. The information describing each granule (the descriptors) would distinguish one granule from another in ways important to the use of the data, such as space-time reference, presence or absence of a given condition or feature, or the value of a parameter.

As examples, an inventory of in situ data might contain information about an individual ship track's sequence of measurements or time sequence of observations taken at a particular site; an inventory of AVHRR satellite data would contain information about individual data acquisitions or passes; and an aircraft data set inventory would contain information about individual scenes or frames within a flight line.

### **Needed Interoperability**

Inventory information held by individual agencies must employ a consistent approach to space and time references to allow cross-inventory searches, given user space-time criteria. In all cases where an internal attribute of a given data set granule is to be available to GCDIS users as a cross-inventory search parameter, it must be consistently named and defined in all applicable cases, using content standards enforced as the metadata are entered. The inventory function should be interoperable with the directory and guide functions; that is, it should allow a user to refer to

summary (directory) or detailed (guide) information about a data set or data set's granules from which were identified by an inventory query as meeting the user's inventory search criteria.

Four levels of agency contribution to GCDIS access system inventory functionality are defined as follows:

- Level 1. Agencies provide access to their online inventory information for inventory searches originating from other agencies' online inventory systems, and support online access to other agencies' inventory systems for their own inventory users. This requires network connectivity for guide sites and network performance level 1 (page 58) for user access and internal GCDIS access system communications supporting the guide function.
- Level 2. Agencies make available online, searchable inventory information using GCDIS access system standard space-time references as applicable and, optionally, GCDIS access system standard descriptors as applicable. This requires network connectivity for guide sites and network performance level 3 (page 59).
- Level 3. Agencies make available in hard-copy form or via a manual service inventory information for priority global change data sets, using GCDIS access system standard space-time references where applicable and, optionally, standard GCDIS access system descriptors.
- Level 4. Agencies make available in hard-copy form or via a manual service (e.g., user service personnel) existing inventories of priority global change data sets, using existing space-time references and/or internal descriptors.

### **Agency Applicability**

The GCDIS access system inventory function applies to all agencies that make global change data and information available to users.

### **Agency Implementation Overview**

Each agency will independently implement its GCDIS access system inventory functionality according to its own approaches and methods. The overviews in Appendix B summarize agency plans.

To establish a framework for individual agency efforts, a cooperative study will be conducted to develop an initial description of the inventory interoperability functional model and system architecture (including functional interface definitions) that will be used by GCDIS access system. This is critical since the scope and

complexity of implementation issues and tasks will be greatly influenced by the chosen architecture. The study will address an inventory content guideline that describes the GCDIS access system standard approach to space-time reference and definition of GCDIS access system standard descriptors (possibly including an initial set, but developing a process for defining them in the future). This study will define GCDIS access system user inventory search capabilities. The initial search capabilities will evolve as users and agencies gain experience). This study will be completed in September 1994. An assessment of other interagency access systems will be included to maximize connectivity and to benefit from the lessons learned by others.

## Browse

### Description

The browse function will allow GCDIS access system users to screen data for usefulness by examining samples of data or summary or reduced resolution versions of large data sets or large data set granules. Examples might include a sample surface meteorological observation, a sample publication of climatological information, a sample ship track of oceanographic observations, or a reduced resolution AVHRR scene (e.g., to check for cloud cover). Browse products can be available online or offline (e.g., printed hard copy or on CD-ROMs). Browse products may be precomputed and available for static browse, access, or may be computed on demand to a users specification dynamic browse.

### Needed Interoperability

The browse function will be data set dependent, but approaches to browsing should be consistent within data types to allow users to compare different data sets. Graphical and image browse products should be designed to facilitate their combined use. Agencies should coordinate in planning browse products for like data held across agencies so that a GCDIS access system user may be able to make meaningful comparisons when browsing similar data available from multiple agencies or agency elements.

The browse function should be accessible from the GCDIS inventory function to allow users to browse granules that otherwise meet users' inventory search criteria or the directory or guide functions (if a data set is most often used as a whole and not described as containing multiple granules).

Four levels of contribution of agency implementation of the browse function are defined as follows:

Level 1. Online, on-demand generation of browse products accessible from GCDIS catalog functions. This requires network performance level 2 (page 59).

Level 2. Online static browse products accessible from GCDIS catalog functions.  
This requires network performance level 2.

Level 3. Offline digital browse products (e.g., on CD-ROMs).

Level 4. Offline hard-copy browse products.

### **Agency Applicability**

The browse function applies to all agencies providing global change data sets of substantial size.

### **Agency Implementation Overview**

Agencies will implement browse functionality individually, but will coordinate to ensure that browse products for similar data sets are similar and support intercomparison between data sets.

## **Order Placement**

### **Description**

The order placement function of the GCDIS access system allows a user, having selected specific global change data or information of interest through use of the GCDIS directory, guide, or inventory functionality, or by being able to specify completely the desired items without the need for use of those services or as a result of previous use of one or more of those services, to place an order for those data and information. The single order may be for a variety of items available from a variety of GCDIS agencies or agency elements. The order must be accessible either independently or from any of the GCDIS access system catalog system functions (directory, guide, or inventory).

Processing and filling an order by the agency or agencies involved will be accomplished agency by agency according to each agency's own procedures and policies. Orders spanning agencies will be filled in parallel by the agencies as if they were independent single orders rather than parts of an overall order for data and information that spans agencies. Standard GCDIS approaches regarding such things as data formats and distribution media are discussed on page 72.

### **Needed Interoperability**

Agencies must provide an interface function for GCDIS access system user order placement at each agency participating in the GCDIS access system. This function must accept user orders, and be able to forward appropriate portions of the order to other agencies for processing.

Three levels of agency contribution to GCDIS access system order placement functionality are defined as follows:

- Level 1. The agency provides an online order placement function that is fully integrated with the catalog system (directory, guide, or inventory) functionality of the GCDIS access system. This requires network connectivity for guide sites and network performance level 3 (page 59) for user access and internal GCDIS access system communications supporting the guide function.
- Level 2. In addition to a manual interface, the agency provides an online order placement function allowing users to order data from the participating GCDIS access system agency elements.
- Level 3. The agency provides a manual interface through user support staff at agency elements offering global change data and information and agreed manual procedures for forwarding orders to other elements of the agency and other agencies elements as needed.

### **Agency Applicability**

This applies to all agencies from which users may order global change data and information.

### **Agency Implementation Overview**

A cooperative study will be done by December 1995 to develop interagency procedures for supporting this function, including determination of the information required from the user to accompany specification of the items ordered.

## **Archive Function**

### **Description**

U.S. global change data policy states that preservation of all data needed for long-term global change research is required. For example, each regional or global observation of the Earth system represents a unique observation of the system, one that will never be repeated exactly; assessment of changes and the mechanisms responsible for them must rely on time series of such unique snapshots. The archives

for global change data must be available for a long time, regardless of the changing interests of the researcher, group, or agency that collected and analyzed the observations. Each GCDIS access system agency has the responsibility to manage, store, and maintain data sets under its purview. A GCDIS access system agency may designate another GCDIS agency to archive some data and information.

*Data set stewardship* is the safe storage and preservation of data sets that are clearly part of the normal operations of each archive site. Each participating GCDIS access system agency will follow government media, storage, and handling standards as prescribed by the National Archive and Records Administration (NARA), the National Institute of Standards and Technology (NIST), and, where appropriate, WDC-A requirements. GCDIS access system archives will be staffed by professionals who understand the data and their sources and processing. Ideally, GCDIS access system archives also will be associated with an agency or university group that undertakes research using the data, verifying quality assessment and documentation of the data. GCDIS access system centers with scientists on the staff or associated with scientist groups are better able to seek out actively and acquire data needed for global change research, but might not be supplied to the archive as a matter of course.

Examples of archive storage and maintenance practices that GCDIS access system archive sites should employ include provision of a clean environment; use of tested and certified media designated for long-term storage; provisions for media backup and off-site storage; adequate security and internal controls to safeguard equipment, software, and media; fire and water protection; and a routine, scheduled media maintenance program, including detection of media deterioration and migration of the data to newer, more stable media when proven reliable.

As part of the archive function, guidelines will be developed for preparing data sets and associated documentation and metadata for long-term permanent retention at GCDIS access system archive sites.

Data-purging policies are the responsibility of the archiving agency. Nonetheless, the GCDIS access system will develop interagency coordination procedures to prevent the loss of important data sets. This will include notification of the IWGDMGC contacts concerning plans to purge a data set at least 1 year in advance of any action to do so, to allow other GCDIS access system agencies to state requirements for the data set or agree to assume responsibility for archiving the data set. NARA procedures will apply if no agreement can be reached on disposition of a data set identified for purging.

### **Needed Interoperability**

All participating GCDIS access system data centers responsible for long-term archiving of global change data sets should adhere to accepted standards for their safe storage and preservation. All centers should follow a common set of cross-agency procedures and guidelines for interagency notification before purging data sets that may be of potential value to global change research.

Four levels of agency contribution to GCDIS access system archive functionality are defined as follows:

- Level 1. As in level 2, with off-site copies (or equivalent).
- Level 2. As in levels 3 and 4 with on-site backup copies (or equivalent).
- Level 3. Full compliance with NARA-NIST guidelines for data storage and preservation.
- Level 4. Storage of data under controlled conditions.

### **Agency Applicability**

GCDIS access system agencies holding global change data sets are responsible for archiving. Some agencies, however, may arrange to have their data sets archived and distributed by other GCDIS access system agencies, either when the data sets are generated or after the data sets have been generated and the originating agency's operational mission (or research program) for those data sets has been completed.

### **Agency Implementation Overview**

An ad hoc archive coordination team will be formed as a near-term interagency implementation activity. The team will develop guidelines by December 1994 for preservation of global change data.

## **Distribution Functions**

### **Description**

The GCDIS distribution function encompasses delivery of data to users on a variety of standard media, including (for moderately sized products) electronic delivery. Standard distribution media for digital data may include 6,250-bpi magnetic tapes, 3480 cartridges, 8-mm and 4-mm cartridges, floppy disks, CD-ROMs, and so forth. Digital data may be distributed electronically via a network, either by allowing online access to the data or by allowing users to request that data be temporarily staged for later access. Nondigital products include publications, charts, and photographic images.

### **Needed Interoperability**

Agencies will agree on a set of standard distribution media for digital data, allowing users a choice, and also allowing users to request data on common media from the GCDIS agencies.

Agencies will agree on a set of standard formats for digital data (whether for offline delivery or online access), including standard distribution formats (for similar data sets), when possible, in addition to native formats that agencies might employ.

Standard formats should be chosen on the basis of acceptance and usefulness to users and availability of support, such as software tools for supporting the formats.

Agencies will agree on distributing documentation (including metadata) with the data. The documentation may be either embedded in data products or distributed as separate hard-copy or electronically accessible text.

Three levels of agency contribution to the GCDIS access system distribution function are defined as follows:

- Level 1. Agencies offer delivery of data (online or offline) in access system standard formats chosen by the user.
- Level 2. Agencies offer electronic delivery of data. This requires network performance level 3 (page 59).
- Level 3. Agencies offer GCDIS access system standard media as a user option.

### **Agency Applicability**

This applies to all GCDIS agencies (or their designees) that provide digital and nondigital data and information products to users.

### **Agency Implementation Overview**

An ad hoc standard distribution format and media team will be formed as a near-term interagency implementation activity. The team will develop guidelines for standard distribution formats and media by December 1994.

## **Billing and Related Administrative Functions**

GCDIS access system billing and administrative functions will be addressed in the future when the agencies agree on the underlying policies.

## Information Functions and Services

### Description

Just as the management of data is handled by major data centers or other organizations in participating agencies, the management of information is handled through centralized scientific and technical programs in a number of GCDIS access system agencies. These programs enable a comprehensive collection of literature produced by the agencies and their contractors and allow aggressive acquisition of documented results of international research and development efforts. The results are compiled into bibliographic data bases for searching and retrieval by the global change community. In terms of U.S. federally funded documentation, the central STI programs maintain systems to provide full archiving and full text retrieval of this material for current and future research. The GCDIS will build upon these existing STI services in Federal agencies.

The interfaces required to provide users with access to bibliographic and text information will be ubiquitous, easy to use, well publicized, and interoperable with the full GCDIS access system. Single-channel access should enable the user to step from data to information or to make decisions about the extent and depth of information desired. Libraries and information centers are experienced in providing access to a broad range of information through catalogs; networks; national and regional systems; and links among public, academic, Federal, nonprofit, and corporate institutions. Decisions concerning levels of access, user fees, summary or full-text information, and links to national and international data bases and networks will be embedded in the overall implementation of the GCDIS access system. These functions will be made as consistent as possible between the data and information providers and communities.

### Needed Interoperability

Information interoperability is needed in terms of systems development, standards in data base definition, and format of data base content.

A new information architecture, whereby a user of global change data and information can sit at a workstation and navigate freely between data files and documentation, is being considered by these STI programs. This work involves systems development, including gateways, navigational tools (e.g., directories), standards development for both information content and for systems, and postprocessing technology for further analysis of the results from the data and information.

Specialized library and information center functions that require interoperability or commonality in the GCDIS include the following:

- *GCMD and library directories and catalogs.* This process will identify components that can interface between the GCMD and comparable directories with library catalogs, data bases, and directories. Front-end software will be developed to assist the end user's search in identifying what resources are available in data centers, libraries, and information centers. A front-end menu will be established to enable the user to step among disciplines, data and information products, and free and fee-based systems.
- *External components.* External components will connect through the Internet or similar networks for access by individual users not necessarily connected to institutions. For example, the general public, students, or staff of small businesses may need answers to crosscutting global change questions. These individuals should not be encumbered by highly technical, scientific findings if all they require is an overview or better understanding of a global change issue.
- *Interactive messaging.* Interactive messaging protocols, practices, and procedures that enable users to relay their unanswered questions or requirements for a different level of assistance will be used. It may be the unserved users who will enable the GCDIS access system to evolve into the most effective system. An easy-to-use messaging system will provide libraries and information distribution centers with orders, requests for additional (perhaps human) intervention, and up-to-date feedback on the types of materials required. The schoolteacher who needs 50 copies of a publication for a class must be routed to a document delivery service that will distribute these materials. Neither a data center nor a library may be the appropriate source.
- *Information clearinghouse.* An information clearinghouse will be established to respond to the increasing and changing information needs of the nonresearcher. Links with the clearinghouse will be automatic for users, whether individual, agency, or institutional. Since thousands of telephone, written, faxed, and other requests are received annually by libraries, inside and outside government, the front-end system should include menu options for large numbers of publications desired by a user. When appropriate, these requests may be handled directly by an agency library or information center or the requests may be directed through the GCDIS access system to another member of the network, such as the clearinghouse. The requester will automatically be notified of the rerouting of the request by the system to avoid confusion and duplication.

### **Agency Implementation Overview**

The following initial steps will be taken through the Library Subgroup:

- Begin systematic efforts to inform librarians and information specialists about the existence, purpose, and activities of the USGCRP, the IWGDMGC, and the Global Change Data and Information Management Program.
- Identify and describe significant library and data collections in the global change area, beginning with the GCDIS agencies, then other Federal collections, then non-Federal (e.g., international, university, nongovernment organizations). Make this information widely available through the GCDIS directory function.
- Begin systematic efforts to identify and evaluate information products designed to aid users of global change information. Recommend the development of such products. Facilitate exchange of existing products between libraries and information centers.
- Complete a strategy for mapping bibliographic information among the DIF, Committee on Scientific and Technical Information, and machine-readable cataloging formats, and make this information widely available to libraries.

## **State and Local Government Access**

State, regional, and local governments may require access to detailed global change information to support decisions at regional scales and to augment their own data collection efforts, especially when intergovernmental coordination is needed. For example, global change information may be used to manage fisheries or to control chlorofluorocarbon emissions.

State, regional, and local governments will be able to access the GCDIS through the Internet and its successors. GCRP agencies will assist them in obtaining the necessary information to facilitate networking access and in obtaining the necessary software to use the GCDIS. Currently, States are able to use some Federal systems through special data-sharing arrangements. These arrangements will be extended to include the full array of GCDIS functionality. Outreach will be necessary through organizations, such as the National Governors Association, to assure that the States and local governments understand the range of services available and the extent and distribution of potentially useful data.

State, regional, and local governments make extensive use of spatial data in their geographic information systems. It is expected that much of the data requested from the GCDIS access system will be spatial data at the highest resolution available. These requests could include the finest resolution available from predictive model runs, as well as remote sensing and in situ monitoring data. Links between the GCDIS and States' geographic data systems are expected.

Historical data may be required to develop status and trends assessments for varying environmental applications. The GCDIS access system will make it easier to obtain this information.

## International Access

In this section a discussion of international access will be followed by a discussion of several international efforts important to global change research in which GCDIS agencies are involved. Additional descriptions of international data and information systems are provided in Chapter 7.

### Description

The agencies acting collectively under the GCDIS access system umbrella will provide international users with access to the same data and information quality and similar delivery services to those available to U.S. users. The agencies will act individually and collectively to secure for U.S. researchers access to internationally available global change data and information services.

Access to global change research data and information anywhere in the world can be ensured more easily by an international extension of the GCDIS access system approach, with the GCDIS access system becoming the U.S. node of an international, distributed, global change data and information network system.

### Needed Interoperability

GCDIS agencies will cooperate in establishing bilateral or multilateral links (including offline and online interoperability of services) with international global-change-related agency, multiagency, and multinational groups and systems. These will include the ICSU-sponsored WDC system (individual elements of GCDIS agencies are or may become ICSU WDCs), the ICSU International Geosphere-Biosphere Program Data and Information System (IGBP-DIS, in which some GCDIS agencies now participate), the Human Dimensions of Global Environmental Change Programme (HDP) being conducted under the auspices of the International Social Science Council (ISSC), and the International Earth Observing System (IEOS) effort being coordinated by the Earth Observation-International Coordination Working Group (EO-ICWG) and the CEOS (in which some GCDIS agencies are now participating). In general, if one or more GCDIS agencies establish links with or interfaces to such efforts or if agency elements become participating nodes in such efforts, then, through the GCDIS access system mechanisms, all GCDIS agencies should gain access and participate. As a result, all users of GCDIS access system services provided by the agencies should benefit from these efforts.

**Agency Applicability**

All agencies that participate in any way with international efforts, such as those mentioned previously, and that produce or hold unclassified or unrestricted global change data and information will ensure that those data and information are accessible through the GCDIS access system to scientists and researchers not only in the United States but also in other countries.

**Agency Implementation Overview**

Each agency will implement its international access functionality according to its own mission and methods. Some agencies with existing international disciplinary or technical relationships have already established extensive links for exchanging global change data and information across the U.S. national boundaries.

## 6. Agency Implementation

### Programmatic Guidance and Milestones

The milestones that were listed in the program plan include USGCRP high-level programmatic milestones to implement the global change data and information management program and a four-phase implementation of the GCDIS. These milestones provide a general framework for the program implementation plan.

#### 1991–1995

- *Support of process studies.* Determine support requirements for bringing sets together in support of process research, particularly field programs and observational campaigns.
- *Support of data fusion and assimilation.* Determine the need for combining data sets across traditional data set boundaries and the strategy for preserving analytic products such as analyzed fields from long-range weather forecasts in anticipation of the creation of integrated global data sets.
- *High quality research data sets.* Establish a coordinated effort by the agencies to identify and set priorities for existing data sets and to develop new global change data sets, *in situ* as well as remotely sensed (including the interagency Pathfinder data sets). This will include continuing the development of paleoclimate and paleoecological data sets needed for global change research, as well as socioeconomic data sets relevant to human dimensions of global change.
- *Model input and output data review.* Determine and set priorities for data sets needed for climate modeling, including existing data sets that may not be widely available. Establish mechanisms to identify and consider issues relating to model-output data.
- *GCDIS phase one.* Initiate a framework for addressing basic issues for content and access; a directory based on, built around, or including the GCMD; prototypes at major data handling agencies; network connections between agencies and to the national research networking infrastructure; and system enhancements within agencies supporting acquisition, processing, quality controlling, cataloging, rescuing, archiving, and retrieving data.

**1996–2000**

- *Integrated satellite and in situ data sets.* Develop integrated, global-scale data sets (including atmosphere, ocean, and land variables) to support diagnostic and predictive models.
- *GCDIS phase two.* Establish full catalog interoperability between data-holding centers for a suite of global change data sets; include links to the GCMD and the guides, inventories, browse systems, and information systems of major data centers from all participating USGCRP agencies.

**2001–2010**

- *Global change historical integrated observation data.* Incorporate those distributed agency holdings of instrumental observations needed for global change research, including historical and contemporary information. This is a culmination of the efforts of earlier years.
- *GCDIS phase three.* Connect major data systems from around the globe with interoperable catalogs and inventories for selected priority global change data sets, enabling researchers to locate data internationally.

**2011–2020**

- *Global paleoclimate, paleoecological data.* Complete the global archive.
- *GCDIS phase four.* Complete full international interoperability between all major data centers and modeling centers for priority global change data sets, including support for high-speed access on an international level.

## Overall Schedule Goals

The overall schedule goals follow from the priorities just outlined and from the milestones presented in the program plan. This presumes formal approval by the agencies of the program implementation plan by January 1, 1994. These are goals only; the ability of the agencies to meet these goals will be limited by mission and budget constraints. Agency-by-agency implementation milestones presented in the functional area subsections that follow represent what the agencies find possible to accomplish within known mission and budget constraints.

The schedule goals will be presented as a set of near-term interagency implementation activities followed by a description of GCDIS phases. The phases represent checkpoint snapshots of GCDIS at 5-year intervals starting with 1995.

### Near-Term Interagency Implementation Activities

These are near-term, first-year (1994) activities that are needed to lay essential groundwork for progressive implementation of GCDIS functionality. These activities will be coordinated by the Content and Access Subgroups and their teams. These 1st-year efforts, once completed, will provide the basis for a major update to this program implementation plan in January 1995.

- Continuation of GCMD operations (directory team);
- Model-output data policy established;
- First inventory of applicable agency data and information completed, and missing archive responsibilities identified;
- Formation of Access Subgroup, User Advisory Group, and teams;
- Agreement on 1994 GCDIS work plan;
- Completion of 1-year lexicon and data dictionary study—analyze requirements and make recommendations to the data community; development of standards with data community for lexicon and data dictionary interoperability with library catalogs and bibliographic data bases (Library Subgroup);
- Beginning of continual effort to identify key data sets in libraries and information centers (Library Subgroup);
- Development of a prototype directory component for the GCMD that summarizes global change information collections in agency libraries and information centers (directory team and Library Subgroup);
- Definition of needed inventory and order interoperability and approaches to progressive implementation (ad hoc agency inventory coordination team);
- Full description by all agencies of their global change data sets and data centers in the GCMD (directory team and agencies);
- Definition of guide, data dictionary, and lexicon interoperability and approaches to implementation (ad hoc guide coordination team);
- Agreement by agencies on guidelines for global change data preservation (ad hoc archive coordination team);
- Agreement by agencies on guidelines for common distribution formats and media (standards team); and
- Definition of the priority set of agency GCDIS components, based on science priorities of global change holdings of the agencies (Access Subgroup with contacts concurrence).

**GCDIS Phase I—Checkpoint 1995**

- Guidance on data and information priorities received from other elements of the USGCRP;
- First inventory of all priority U.S. data completed;
- Policy and process for peer review of data sets established;
- First regular climate status-and-change summary for North America produced;
- Review of implementation by the CGED;
- Major update of program implementation plan, with high priority for focusing implementation efforts on priority data sets, services, and the agency elements involved (including agency-specific milestones for meeting content objectives);
- Initial level of interagency coordination of user support services (e.g., user referral);
- Prototype implementations of inventory functions across a set of agencies;
- Prototype implementation of multiple-server configuration of GCMD;
- Plan for implementation by 1997 of prototype guide function across a set of agencies;
- Completion of a study and plan for implementation of the GCDIS order placement function; and
- Implementation of common media for distribution of digital data.

**GCDIS Phase II—Checkpoint 2000**

- Recommendations on priorities received from the NAS (1996);
- First inventory of all priority data and information internationally completed (1996);
- Individual data and information set priorities established for the total USGCRP (1996);
- First regular climate status-and-change summary for entire Earth produced (1996);
- Highest level of service for all essential data and information is available (1997);
- Review of implementation by CGED (1997);
- Continual, active, full coordination of agency user support functions (1997);

- Implementation of the order placement function at a set of agencies, including prototypes of an online function (1997);
- A set of agencies offer common formats and media for priority global change data sets (1997);
- Implement preservation standards and practices for rescue and preservation of historic information (pre-1950) in library and information center collections (1997);
- Implement electronic access to selected scientific information data bases at agency scientific/technical information distribution centers (1997);
- Design and prototype (with users) a front end for library catalogs and data bases that will provide a preview capability for global change information sources (1997);
- Prototype interoperability of GCMD and selected agency library catalogs (1997);
- Complete extension of the directory function to include data relevant to human dimensions of global change (1998);
- Next-to-highest level of service for all high-priority data and information is available (1998);
- All agencies compliant with data preservation guidelines (2000);
- Full implementation of catalog system functions across all agencies (as applicable for agency global change mission) (2000); and
- Complete digitizing of library publications containing significant data sets or metadata held in agencies (2000).

## **Agency Implementation**

The agencies will establish version alpha of the GCDIS consisting of their components of the GCDIS that will be available on April 1, 1994. Version alpha will be the base on which the GCDIS will grow and evolve. Agency summaries of their portions of version alpha follow. The version alpha components for each agency and the additional components that will be part of the GCDIS by April 1 of 1995 and 1996 are listed in Appendix B. Because of the evolving nature of the USGCRP and of the GCDIS, detailed projections beyond 1996 have not been given—they will be included in future updates of this implementation plan.

### **Department of Commerce**

NOAA is the primary agency within the Department of Commerce (DOC) for GCDIS participation as both supplier and user of climate and global change data and information products. Other DOC agencies will participate in the GCDIS as their roles are better defined and as the system evolves. For example, the U.S. Bureau of the Census, through its Center for International Research, has initiated global-change, data-related, collaborative activities with the CIESIN. The Census Bureau is also attempting to increase public awareness of and use of its own data through the development of its own Internet site and other online electronic collaboration with appropriate parties.

NOAA's three National Data Centers, and the NOAA Satellite Active Archive, the NOAA Directory Services, the NOAA Library, and the NOAA Network Information Center (NIC) are agency components that will participate in GCDIS operations. As the GCDIS evolves, NOAA, through its NOAA-wide Data System Modernization, will add additional capabilities and improved interoperability for the GCDIS user community. NOAA's Data System Modernization is now in the planning stage; system implementation is expected to start in 1995 and continue through at least 2004.

### **Department of Defense**

The DOD routinely collects environmental data globally in support of DOD operations. The DOD also conducts mission-related research into environmental processes and conditions that affect defense operations, tactics, and systems. The DOD does not have a mission requirement to archive data for civil uses. DOD data products are made available to the appropriate U.S. national archives for subsequent use by the GCDIS.

### **Department of Energy**

The Atmospheric Radiation Measurement (ARM) program is a DOE-sponsored, global-change-research effort designed to improve the modeling of cloud radiative forcing in general circulation models. The primary user community for ARM data is the atmospheric research community; secondary users are the broader scientific community that has interest in some of the meteorological or radiative measurements that the ARM can provide.

The ARM Archive receives and manages a broad variety of data, such as

- Measurements and observations of radiative and atmospheric phenomena,

- Information and documentation about the data streams that allow scientists to interpret the data,
- Information about the instruments producing the data, and
- General information about the ARM project.

This information comes from a variety of sources within the ARM project and from several external sources. The primary data source is the ARM Cloud and Radiative Testbed (CART), which will consist of three highly instrumented sites located worldwide. The first site, centered close to Lamont, Oklahoma, has been taking data since June 1992. Data fusion products and data quality measurement products are generated at the ARM experiment center using data from the CART sites and observations from external sources such as satellites. Information about instruments, data quality, and instrument operations comes from instrument developers, instrument mentors, site operations staff, and scientists.

### **Department of the Interior**

As the major Federal land manager and the primary Federal agency responsible for managing the Nation's natural ecosystems, fish and wildlife, and energy and water resources, the Department of the Interior (DOI) is particularly concerned about the potential short- and long-term effects of climate and other environmental change on these lands and resources. The DOI's global change research is addressing topics such as hydrologic and geologic processes and resources, land use, land cover, biological habitats, resources, and diversity; past global change recorded in the physical, chemical, and biological record; land surface and solid Earth processes that relate to environmental change; geography and cartography; polar and arid region processes; ecosystem modeling and dynamics; and resource ethnology.

The DOI bureaus collect, maintain, analyze, and interpret short- and long-term land, water, air, biological, and other natural resource data and information in support of their missions. These efforts have always included maintenance of high-quality, long-term data sets, including cartographic, land cover, geologic, hydrologic, ecological, and biological data from both satellite- and aircraft-based remote sensing and terrestrial-based observations. The DOI will provide access to directory-level descriptions of these data (and inventory-level information where feasible) through DOI and interagency directories and clearinghouse mechanisms such as the National Geospatial Data Clearinghouse. The USGS Global Land Information System (GLIS) will be one of the primary inventory-level interfaces with the GCDIS and an access point to DOI global change data. The DOI also participates in the NASA Earth Observing System (EOS) program through the Earth Observing System Data and Information System (EOSDIS) Land Processes Distributed Active Archive Center (DAAC) at the USGS Earth Resources Observations Systems (EROS) Data Center

(EDC), where capabilities are being developed to process, archive, and provide online information system access to EOS land-related data sets, such as those from the Moderate Resolution Imaging Spectrometer (MODIS).

### **Environmental Protection Agency**

The EPA has the primary responsibility for data on environmental quality and the distribution and effects of pollutants on human and ecological health. As such, it is both a supplier and consumer of information on the environment, with a large potential for beneficial interchange with the GCDIS that goes well beyond its current role in the USGCRP.

Data and information for most of EPA's programs are stored centrally at the EPA National Computer Center in Research Triangle Park, North Carolina. The EPA *Information Systems Inventory and Access EPA* are key documents that are available in both hard-copy and electronic versions describing the agency's holdings. The EPA is attempting to increase public access to its holdings through the development of an Envirofacts data base, GCDIS-compatible access tools, and a public access server at Research Triangle Park.

Data and information from the focused portion of the USGCRP (such as the North American Landscape Characterization data set) will initially be made available through arrangements with other GCDIS archives. Contributing data and information from EPA's regulatory and scientific programs, not derived from the USGCRP activities, will be incorporated into the Envirofacts data base as funding becomes available and requirements are known from the USGCRP.

### **National Aeronautics and Space Administration**

NASA provides access to Earth science data through several discipline-specific data centers and data systems. Most of these can be accessed automatically from the GCMD. These centers and systems provide various levels of service for data processing, distribution, and archiving. The numerous disciplines supported include climate, oceanography, land science, hydrology, biogeochemical dynamics, sea ice, geophysics, atmospheric dynamics, radiation budget, and human dimensions.

Researchers may access most of the systems through the Internet and dial-up lines, or they may visit hard-copy browse facilities. Data are delivered electronically or on standard media, such as 9-track magnetic tapes, 8-mm cartridges, or CD-ROMs. System capabilities allow users to search, locate, select, and order products. Searches usually can be limited by geographic area, time, or geophysical parameter. Electronic

access is generally free to research users, but a fee may be charged to cover the marginal cost of filling the requests.

### **National Science Foundation**

The National Science Foundation (NSF) sponsors a large and diverse research community that both uses and produces global change data and information. Although the NSF has no formal responsibilities for archiving and distributing data and information, it supports a major facility for meteorological, oceanographic, and climatology data sets at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. By and large, however, NSF-supported scientists rely on other Federal agencies for much of their data and information needs and, when appropriate, for archiving and disseminating the research products they produce.

The NSF expects its supported investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections, and other supporting materials created or gathered in the course of the research project. Presently, the enforcement of this requirement varies considerably across the agency. For example, social and behavioral data sets are deposited in an archive for distribution within a year after the completion of a grant. Other discipline divisions at the NSF have no formal policies. The NSF will implement, in collaboration with other GCDIS agencies, a process by which important global change data sets produced with NSF support will be archived, managed, and disseminated for broad community use. Deciding which products are appropriate for this treatment and how the activity will be funded will involve a multilateral process among the research principal investigators, the sponsoring agency program manager, and the appropriate data center and its sponsoring agency. A general rule of thumb for funding responsibility is that the agency program manager should support those activities that are required for the research project itself, and the data center should support those that are required to serve the broader community needs.

### **U.S. Department of Agriculture**

The U.S. Department of Agriculture (USDA) has initiated several new projects to support the need to make global-change-related data and information readily available to its various user communities. This inventory includes special information required to define spatial data and models and the particular data sets required for operation of the models. Data bases inventoried are being categorized as weather data (e.g., temperature, precipitation), atmospheric data (trace gases, deposition), soils data, forest data, plant and vegetation data, animal data, pest data, hydrologic data, economic data, and more. An automated locator and directory system to house the inventory metadata is being made and soon is expected to be interoperable with the

GCMD and other servers, such as the WAIS servers installed at other Federal locations.

The National Agricultural Library will continue in its major role in identifying, cataloging, and providing access to worldwide published information related to global change issues. The Current Research Information System will be available through the Internet and commercial information distribution systems, and will provide information on the status of research projects funded by the Cooperative State Research Service of the USDA.

## **The Evolution of the GCDIS**

It is abundantly clear that the GCDIS must be built so that it can evolve. This need is driven by a number of factors. First, the USGCRP program, which the GCDIS serves, is even now changing in terms not only of its scope of research, but in terms of the relative priorities of individual research areas and of the increasing importance of assessments that have policy relevance. For example, the USGCRP focus has recently been expanded from that of climate change to also include stratospheric ozone depletion, biological diversity, forests, and desertification. Such changes will continue as the USGCRP itself evolves.

Second, from the time when data management was barely an afterthought for many programs and projects, its importance has rapidly grown in nearly all areas of endeavor. This very growth, coupled with the broad scope of the USGCRP, means that the GCDIS must evolve to stay compatible with developments in a wide range of fields over which it has almost no control. Examples are Federal geographic standards development, government information locator systems, library search-and-retrieval standards, international systems, health- and other human-dimension-related data base developments.

Third, the technology which the GCDIS will have to use to meet its users' needs is rapidly changing. This change is increasingly in response to economic imperatives and technical needs largely outside the USGCRP. Examples include developments in such major areas as high-performance computing, data superhighways, and commercial communications and cable systems.

To be able to evolve as necessary, the GCDIS will do the following:

- Plan the system so that it is capable of interfacing with other systems having a wide range of capabilities. The GCDIS will be inclusive as opposed to exclusive in terms of only being able to serve users in narrowly defined modes. The levels of service defined for access in Chapter 5 are an example of this approach.

- Establish working relationships with other groups, such as the FGDC and the library community, to help guide the evolution.
- Implement the user coordination system described on page 14 to provide user guidance as to necessary changes.
- Conduct pilot demonstrations in areas where interfaces with new user communities are needed, such as the library community, or new technical capabilities, such as the NREN.

As an example of evolution in the program, assessments and policy questions will increasingly require data and information from the biological, social, and economic sciences. These biological and human dimension data sets will largely result from investigations funded externally to the USGCRP. They will range from process research results to operational modeling and monitoring and will represent basic science, as well as regulatory and resource management information. Considerable local as well as Federal data and information are anticipated. As the GCDIS evolves, it will be increasingly challenged to develop the necessary infrastructure (both technical and political) to access and use these disparate data and information in USGCRP assessments.

Initially, elements of socioeconomic data will be held in the Socioeconomic Data Center (SEDAC) funded through the EOSDIS. The SEDAC will be part of the EOSDIS network, and its data and information will be available through the EOSDIS. In addition to the data and information that the SEDAC will hold directly, an Information Cooperative has been organized by CIESIN to arrange for access to data and information held by institutions throughout the world. The SEDAC will also provide pointers to these relevant data sets. This follows the general GCDIS model of a responsible data center for a particular type of information and a uniform policy for data distribution and charging.

In addition to the SEDAC and CIESIN Information Cooperative, USGCRP member agencies will directly provide important biological and socioeconomic data and information, of appropriate granularity, through their individual implementations of the GCDIS. For example, programs such as the EPA Environmental Monitoring and Assessment Program, the DOI Biological Survey, and the NOAA Coastwatch will develop important data and information on biological resources that will have applicability to USGCRP analyses. Including these data and information in the GCDIS will proceed in an evolutionary manner.

These programs are typical of many Federal and State programs outside the USGCRP that have data needed by the USGCRP. Being outside, many have very limited funding for making such data available for USGCRP use and also may require such special administrative arrangements as memoranda of understanding. The need for developing GCDIS access to such data sets that were not developed by the USGCRP itself is identified as a special issue in Appendix A.

## **7. International Links and Coordination**

Initially, GCDIS links and access to international data holders, providers, and users will be accomplished through existing agency participation in the ICSU WDC system. A number of U.S. data centers associated with implementation of the GCDIS currently have responsibility under the WDC system for a broad range of important climate and global change data. All GCDIS archive sites will pursue WDC status through the ICSU or be linked with existing WDCs. The links for exchanging and transferring global change data and information among WDCs vary from magnetic tapes and cartridges and CD-ROMs to data communications networks.

GCDIS international links among WDCs will be completed or supplemented by other media or networks to transfer Earth observation data and information from satellite and in situ sources for other operational or noncommercial uses that benefit the public. Those existing and emerging international systems, services, and their coordinating mechanisms are described in subsequent sections. Expanding links to other international organizations and programs concerned with global change data management will be useful in several ways to implementation of the GCDIS.

First, links to data centers and archives outside the United States will be necessary to access data relevant to the USGCRP and to make U.S. data and information available to collaborating researchers and governments abroad. To be useful, these must be permanent archives, so that the data will be readily accessible to users well into the future. This is particularly true for global change data, inasmuch as assessment of change and the mechanisms responsible for it must rely on long time series that are irreproducible. Archives for global change data must therefore be maintained on a long-term basis, irrespective of the changing interests of the scientist, group, or even agency that collected and analyzed the observations. Several existing internationally sponsored data management and archive systems, to be described briefly later, have the long-term commitment needed to serve as global change archives.

Second, many of these international institutions are associated with an agency or university group that undertakes research using such data and are therefore more than just data libraries. The personnel associated with such centers perform the highly useful function of using the data themselves, and therefore perform quality assessment and ensure good documentation of the data. Such centers are in good position to seek out actively and acquire data that are needed for global change research, but that may not be supplied to the archive as a matter of course.

Third, timely international agreement regarding standards, protocols, and formats for new types of global change data and information will facilitate access and exchange of global and regional data sets and research information. Several existing international technical forums for developing such agreement are described later in this chapter.

Fourth, agreement on the part of international partners—particularly governments—will be necessary to overcome policy-related impediments to full and open exchange of data and information. Intergovernmental organizations provide a useful forum for discussion and agreement among governments on data management and exchange principles, policies, and practice.

Major impediments to the full and open exchange and distribution of international data for global change research include

- Lack of a comprehensive worldwide inventory of major existing data bases;
- Major uncertainties about the quality and accuracy of existing data sets;
- Lack of consistent international policies and protocols for data exchange, particularly with regard to charging;
- Unnecessary or inappropriate (overclassification) restrictions on data access on the part of some governments; and
- Lack of long-term (10 to 20 years) commitment from any national or international organization to develop policies and support programs to facilitate data archiving and exchange in some new areas of global change research.

It should be noted that the international community is beginning to recognize the need for consistent policies for international data exchange. Several international organizations, including the Intergovernmental Oceanographic Commission (IOC), the IGBP, and the CEOS, have formally adopted policy statements consistent with the principle of full and open exchange of scientific data and information. Although not binding on member countries, they constitute a first step toward establishing an international norm. The United States has adopted a similar policy—the *U.S. Data Policy Statements for Global Change Research* (see Table 1).

## **Existing International Data and Information Systems and Services**

Some of the existing systems that may serve as GCDIS archives or that may in other ways be part of the data and information dissemination system needed for global change research will be described.

### The ICSU World Data Center System

The ICSU WDCs were originally organized for the International Geophysical Year (1957–58) and, upon recommendation of the ICSU, have continued since as a network linking data providers to data users around the world. This system is operated by volunteering national organizations on agreed principles and policies contained in the ICSU's *Guide to the World Data Center System*. All the WDCs are staffed, funded, and maintained exclusively by the countries in which they are located. The four main activities of the WDCs involve acquisition, exchange, dissemination, and archiving of solar-geophysical and related environmental data and information. As already noted, the WDC system will provide the initial GCDIS mechanism for international access.

The WDC system now consists of 44 archive and data centers in several countries: The United States, Russia, China, Czechoslovakia, the United Kingdom, France, Belgium, Denmark, Japan, and India. WDC–A is hosted by the United States, and is sponsored by NASA, NOAA, the USGS, and the U.S. Naval Observatory. Oversight is provided by the NAS CGED, which maintains at the NAS the WDC–A Coordination Office. The individual U.S. centers are organized and funded by the parent agencies, but the NAS CGED is the information/organizational channel that connects the WDC–A system to ICSU's supervisory/advisory group, the Panel on World Data Centers.

In the United States, USGCRP agencies now operate 12 WDC–As for glaciology (snow and ice), marine geology and geophysics, meteorology, oceanography, rockets and satellites, rotation of the Earth, seismology, solar-terrestrial physics, solid Earth geophysics, paleoclimatology, land cover, and atmospheric trace gases. All of these WDC–As are collocated with national disciplinary data centers throughout the United States. The WDC–As will provide a fundamental GCDIS mechanism for international access and data exchange among corresponding WDCs in other parts of the world.

One of the weaknesses of the ICSU WDC system is that there are no specific recommendations for archiving and exchange of satellite-based data that are essential for global change studies. Whereas some of the conventional data centers (e.g., oceanography, meteorology, snow and ice, the new center for remotely sensed land data) do receive and archive data products derived from satellite observations, the present configuration may not meet all needs of the GCDIS. However, the present system (WDC–A) maintains links with satellite data systems inasmuch as three of the WDC–A centers are in agencies or groups that also house NASA EOS DAACs, namely the USGS EDC in Sioux Falls, the DOE's CDIAC in Oak Ridge, and the University of Colorado's National Snow and Ice Data Center in Boulder. To further expand the connection between WDC–A and satellite data systems, further

discussions are underway with NASA regarding four additional DAAC-related data centers at the Goddard, Langley, and Marshall Space Flight Centers, and the Jet Propulsion Laboratory.

### **The World Meteorological Organization Data System**

Under the aegis of the WMO, several countries operate data centers closely related to the WCRP and IGBP–Global Change: The World Ozone Center at Toronto; the Global River Runoff Data Center at Koblenz, Germany; the WDC for GHG, Tokyo; and the Global Precipitation Center, Germany. A less active center that is being revived for WCRP and global change studies, and integrated again into the WMO system, is the International Radiation Center in St. Petersburg, Russia.

### **The Intergovernmental Oceanographic Commission**

The Intergovernmental Oceanographic Commission's (IOC) International Oceanographic Data and Information Exchange system (IODE) consists of about 45 National Oceanographic Data Centers as well as seven specialized data centers for specific ocean observing systems and techniques. The three ICSU WDC's for oceanography, located in the United States, Russia, and China, form the apex of this pyramidal system, serving as the final archive and distribution centers for oceanographic data products once they have been processed and quality assessed through the other parts of the IODE system. The successful merging of the IOC and ICSU data center systems for oceanography may serve as a useful model for establishing relationships between the GCDIS and the ISCU WDC system.

### **The Federation of Astronomical and Geophysical Services**

This ICSU system, sponsored by the International Union of Geodesy and Geophysics and the International Union of Radio Science, is also an outgrowth of the International Geophysical Year. The Federation of Astronomical and Geophysical Services centers that relate directly to the GCDIS are the Permanent Service for Mean Sea Level in the United Kingdom, the World Glacier Monitoring Service in Switzerland, and the Sunspot Index Data Center in Belgium.

### **Other Data Services**

There may be other independent geophysical data services, but most of them relate to solid-Earth, upper-atmosphere, or space disciplines that have indirect relation to global change studies.

## **International Mechanisms for Planning, Coordinating, and Implementing**

### **The International Geosphere–Biosphere Program and the IGBP–DIS**

The IGBP, established by the ICSU, is designed to achieve an understanding of Earth and its environment, to improve our ability to detect global change and plan intelligent responses. Previous ICSU programs (concerning the lithosphere, hydrosphere, cryosphere, atmosphere, biosphere, magnetosphere, and ionosphere, for example) have successfully studied components of the Earth system. The intent of the IGBP is to work toward realizing a synthesis on a global scale of how these individual components interact to produce the dynamics of the whole Earth system, with the goal of predicting the future state of the planet into the next century.

In addition to the ICSU WDC's, which either act as archive and distribution centers, or which produce special data products for research, there is another ICSU data activity important to the evolution of the GCDIS. Detailed planning for IGBP data management has started under the Working Group on Data and Information Systems for the IGBP (IGBP–DIS). Based in Paris and supported by NASA and the French Ministry of Sciences, its overall goal is to improve the supply and management of data and information that will allow the IGBP to satisfy its agenda.

The IGBP–DIS has been most successful as the focal point for IGBP Core Projects, to identify data needs, to build specifications for missing or presently inadequate data sets, to conduct IGBP data workshops, and to serve as an information link between such services as the GCMD and the user community. Four initial projects involving data or metadata have been identified in advance of the specification of data needs by Core Projects:

- Land-Cover Change Pilot Project,
- Vegetation Imagery Diskette Pilot Project,
- Overall systems analysis of important data sets for global change studies, and
- Directory of national or disciplinary directories of data sets useful for global change studies.

These pilot projects have proved to be very useful. For example, as an adjunct to the Land-Cover Change Pilot Project, the IGBP–DIS was instrumental in defining the needs and specifications for the AVHRR 1-km global pilot data set now being implemented by the USGS, NASA, and NOAA. A CD-ROM containing results from the 1st year will be issued during 1993. The IGBP–DIS also is sponsoring an investigation of the possibility of obtaining useful land surface temperature data from

satellite observations, and was the initial sponsor of the IGBP African Diskette Pilot Projects, the successful implementation of which resulted in the NOAA-EPA Global Ecosystem Database (on CD-ROMs) and the African Diskette Educational Project. Additionally, a new global soils data base has been started.

USGCRP agencies operating WDC-As have begun to participate in the IGBP-DIS diskette pilot projects. Those agencies' GCDIS plans need to be modified to include producing and disseminating space-based and in situ regional and global data sets on diskettes or other media to facilitate international access to global change research data and information.

### **The Human Dimensions of Global Environmental Change Programme**

The HDP, established in 1990 under the auspices of the ISSC (the social-science counterpart to the ICSU) is initiating an international program of research to address the human dimensions of global environmental change. The HDP is defining its program, analyzing data and information system requirements, and initiating focused projects.

Seven broad topical areas of research have been identified: social dimensions of resource use; perception and assessment of global environmental conditions and change; effects of local, national, and international social, economic, and political structures and institutions; land use; energy production and consumption; industrial growth; and environmental security and sustainable development.

Focused projects include Land Use/Land Cover Change (together with IGBP), and the Global Omnibus Environmental Survey. Working groups are being formed to address focused projects in other areas of research.

The HDP Data and Information System (HDP-DIS) will be developed through collaboration with extant social science archives and information sources by forming a data- and information-sharing network that will ultimately interoperate with natural science data and information system initiatives such as the IGBP-DIS. The functions of HDP-DIS include

- Developing science data and information resource requirements for HDP focused programs and determining their availability and sources;
- Providing services in support of HDP science initiatives, to include data resource discovery, documentation (cataloging) and archiving; and maintaining information services such as guides and bulletin boards;
- Identifying computer-processing and data-resource services in support of HDP science computing, analysis, and data integration needs such as

geographical information systems, decision support and modeling capabilities; and

- Accessing extant networks and developing interoperability among extant and future information archives and centers that will support HDP science initiatives.

HDP-DIS is now undertaking an assessment of requirements and implementing elements of an emerging concept of operations. An HDP-DIS Data Users Group and two advisory panels are being formed. These are a Data Resources Panel, chartered to determine the requirements of the HDP focused programs, and a Data Sharing Panel that will develop working relationships among active archives and sources for relevant and required socioeconomic data.

The development of an Internet-based HDP information service is being explored. A catalog interoperability test will explore the applicability of the CEOS metadata model and the IDN concept in the HDP domain. This metadata-sharing experiment will use socioeconomic data archives and sources worldwide in one of the first attempts to link data resources using network-based resource discovery tools. Coordination with the CEOS-IDN is planned to ensure future interoperability. Coordination with the System for Analysis and Training is essential to ensure technology access for developing countries. HDP-DIS activities have been integrated since inception with NASA's EOSDIS through the EOSDIS Socioeconomic Data and Applications Center, managed by CIESIN, which is providing initial HDP data services and coordinating system development.

### **The Committee on Earth Observations Satellites**

The CEOS was created in 1984 as a result of the Economic Summit of Industrialized Nations to serve as a focal point for informal international coordination of space-related Earth observation activities. Its areas of concern are mission planning, data access, networking, data product standards, compatibility and interoperability of data products, services (including catalog, data access, etc.) and applications, as well as instrument calibration and validation of satellite data with ground truth data. The CEOS is a voluntary mechanism operating by consensus.

The members of the CEOS are NASA and NOAA for the United States, the Canadian Space Agency (CSA) for Canada, the Chinese Academy of Space Technology and the National Remote Sensing Centre for China, the Centre National d'Etudes Spatiales for France, the Indian Space Research Organisation for India, the Instituto Nacional de Pesquisas Espaciais for Brazil, the Science and Technology Agency (STA) for Japan, the British National Space Centre for the United Kingdom, the Agenzia Spaziale Italiana for Italy, the Deutsche Agentur für Raumfahrt-

Angelegenheiten for Germany, the Commonwealth Scientific and Industrial Research Organization for Australia, the Russian Space Agency and the Russian Federal Service for Hydrometeorology and Environment Monitoring for Russia, the European Space Agency (ESA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) for Europe, the Swedish National Space Board for Sweden, and the National Space Agency of Ukraine. In addition, CEOS observers include the Norwegian Space Centre for Norway, the Canadian Centre for Remote Sensing for Canada, the Science Policy Office for Belgium, the Crown Research Institute for New Zealand, and the Directorate General 23 for the European Community.

NOAA and NASA have worked for the past several years to broaden the perspective of CEOS beyond its original narrow focus of land-oriented applications of remotely sensed data to encompass the broader range of global change research and environmental monitoring. The CEOS Plenary group has accepted this view of CEOS, and has accepted as CEOS affiliates the Food and Agriculture Organization (FAO), the ICSU, the IGBP, the IOC, the WCRP, the WMO, the UNEP, the Global Climate Observing System (GCOS), and the Global Ocean Observing System (GOOS).

The CEOS comprises a high-level plenary group with a policy focus and two supporting working groups. The Sensor Calibration and Performance Validation Working Group is now focusing on sensor intercalibration required for global change research. The Working Group on Data (CEOS-WGD) is concerned with coordination of data management. The CEOS-WGD is active in the development of the CEOS IDN as a first step toward international data and information system interoperability, networking, development and adoption of standards for international data exchange formats, development of a lexicon and data dictionary, and assembly of a global 1-km AVHRR data set.

The WGD has developed a coordinating relationship with Panel 2 of the Consultative Committee on Space Data Systems, which is concerned with developing (through a formal process) internationally agreed recommendations for standards in areas relating to data and information exchange.

The CEOS-WGD has created subgroups working at an implementation-coordination level in several areas. These are a Formats Subgroup, working on data exchange formats; a Catalog Subgroup, working on the CEOS-IDN and catalog interoperability in general; a Network Subgroup, working on ground networking between CEOS agencies internationally; and a subgroup on Auxiliary Data Sets.

### **The Earth Observation–International Coordination Working Group and the International Earth Observing System**

The objectives of the EO–ICWG are to promote the IEOS as an integrated Earth observing system to advance understanding of the Earth system, to promote effective use of Earth observation spacecraft (e.g., by coordinating payload planning), and to promote continuity of operational services provided currently by NOAA’s polar-orbiting satellites and development of future operational services.

The EO–ICWG was formed in 1986 by NASA, NOAA, and the ESA to coordinate polar platform programs and payload planning. Current participants are: NASA and NOAA for the United States; the ESA and the EUMETSAT for Europe; the STA, the Ministry of International Trade and Industry (MITI), the Japanese Environment Agency (JEA), the Japanese Meteorological Agency, and the National Space Development Agency (NASDA) for Japan; and the CSA for Canada. These agencies are known collectively as the International Partners. The current role of the EO–ICWG is to coordinate and implement the IEOS, which is seen as the coordinated aggregate of the participating agencies’ end-to-end Earth observing systems (i.e., space and consensus).

The EO–ICWG has to date identified the following missions and their supporting ground systems as elements of the IEOS: for the United States, NASA’s EOS missions and NOAA’s operational Polar-Orbiting Operational Environmental Satellite series; for Japan, NASDA’s Tropical Rainfall Measuring Mission (TRMM) and Advanced Earth Observing System (ADEOS) missions; and for Europe, the ESA Polar-Orbit Earth Observation Mission (POEM) missions. In many cases other agencies will provide instruments, including the MITI for EOS, the CSA for EOS, the MITI and NASA for ADEOS, NASA for TRMM, and NOAA and EUMETSAT for POEM.

The EO–ICWG has recognized that implementation of the IEOS would require coordinated efforts in data management. Rather than form a new data management group, the EO–ICWG requested that the CEOS, which had already created an effective forum for informal international coordination, assume responsibility for addressing IEOS data management issues.

### **Proposed Global Climate Observing System and Global Ocean Observing System**

The GCOS and the GOOS are two newly planned and interrelated global-scale observing systems with significant data management components. When even partially implemented, it will be essential for the GCDIS to establish links to its elements. This should not be difficult, since U.S. agencies—particularly NOAA—are

playing key roles in the planning and implementation process. Internationally, the EOS is expected to provide a core component of the mature GCOS and GOOS systems. Related data management responsibilities have not been specified, but no doubt the WDC system will play a key role, at least in the interim.

Planning for the GCOS is led by the WMO, with joint sponsorship by the UNEP, the IOC, and the ICSU. The GCOS will collect, manage, and distribute data and information relevant to prediction of climate variability and global climate change, including in situ and remotely sensed data. Four components are planned: oceans, atmosphere, terrestrial, and cryosphere. The WDC system is studying how various relevant centers can undertake data management roles for data resulting from the GCOS and the GOOS, bearing in mind that the data flow is primarily designed for operational (near-real-time) use, and special arrangements may have to be made for parallel or delayed processing for research-grade data products.

Planning for the GOOS is led by the IOC, with joint sponsorship by the WMO, the UNEP, and the ICSU. The GOOS will consist of five modules—climate change, health of the ocean, living marine resources, oceanic conditions (for operations), and coastal zone management. The climate change module of the GOOS will, in large part, serve as the ocean component of the GCOS. Initially, the GOOS will be based on existing programs, including, for example, World Weather Watch, Voluntary Observing Ships, the Integrated Ocean Services Station System, and the IODE. Strong links are planned with such research programs as WOCE, the Tropical Ocean Global Atmosphere Experiment (TOGA), and the Joint Global Ocean Flux Study (JGOFS), and the GOOS will no doubt be modified by their results.

Both the GCOS and the GOOS were strongly endorsed by delegations to the U.N. Conference on Environment and Development, held in June 1992 in Rio de Janeiro, and thus have broad support among both industrialized and developing nations. Access to relevant data and information on climate variability and global change is considered essential for developing countries to wisely manage their coastal areas, water resources, agriculture and fisheries, and land use for sustainable development.

### **The Intergovernmental Panel on Climate Change**

The IPCC was established to advise the International Negotiating Committee for the Framework Convention on Climate Change, and is sponsored by the WMO and the UNEP. The IPCC conducts a scientific assessment of climate change every few years, focusing on immediate issues and consequences, and prepares in-depth summary information for policymakers, so that realistic response strategies can be created to manage climate change issues. Thus, the panel is both a user and a synthesizer or producer of information.

Many of the requirements specified by the IPCC are being addressed by continuing international programs. Relevant international projects are coordinated by the WCRP, sponsored by the WMO, the IOC, and the ICSU, and by the IGBP, which is also sponsored by ICSU. The program on the Human Dimensions of Global Environmental Change, recently established by the ISSC, is expected to fulfill some of the IPCC requirements for social science data and information.

---

---

# **Appendices**

## A. Data Capture

### Data and Information Sources

To achieve its objectives, the USGCRP requires a wide variety of data and information from many disciplines for long periods of time. The data and information exist in both digital and nondigital forms and include raw data from observation systems and surveys; value-added data from data assembly activities; derived data and information from models and other investigations; long-term as well as short-

**Table A1. Data and Information Agency Sources**

Agency	Greenhouse Gases, Aerosols, and Ozone	Water, Energy, and Sea Level	Ecological Systems	Human Interactions
DOC	F, X	F, X	F, X	F, X
DOD	F, X	F, X	X	X
DOE	F, X	F, X	F, X	F, X
DOI	F, X	F, X	F, X	F, X
EPA	X	F, X	F, X	X
HHS				X
NASA	F	F	F	F
NSF	F, X	F, X	F, X	F, X
Smithsonian	F	F	F, X	F, X
State Dept. (international)	X	X	X	X
TVA		F, X		
USDA	F, X	F, X	F, X	X

*F* indicates programs that have the USGCRP as their primary objective;  
*X* indicates those that do not.

term data and information sets; historical, current, and future data and information; in situ and remote observations; information from other studies; and references to data and information that are produced outside the USGCRP, such as major international and national programs.

Extensive collections of such data and information critical for global change research are now supported in the Federal agencies and many other sources such as libraries, State and local governments, university and other research efforts, and the international community. Within the USGCRP, the *focused* program includes data and information from those programs identified by the agencies as having the needs of the USGCRP as their primary objective. A wealth of other data and information, gathered for purposes other than global change research, also is critical for the USGCRP, including those identified in the USGCRP *contributing* programs. Some of these data and information may not reside in organized data and information systems, and are not easily accessible or well documented. Table A1 shows agency sources for data and information critical for the USGCRP arranged by its central priorities. *F* indicates programs that have the USGCRP as their primary objective, and *X* indicates those that do not.

## The Special Issue

The USGCRP recognized at its inception that a substantial part of the data and information needed for global change research would not be created by the focused programs created specifically for USGCRP purposes. The Federal Government supports thousands of individual research projects in Earth, environmental, and human sciences that are not part of the USGCRP focused programs. These projects have produced such national treasures as the archives held in the NDCs, the daily data collection and analyses provided by the National Weather Service, university-based holdings of tree ring samples and ice cores, forest inventories from the USDA Forest Service, cartographic data and stream flow records from the USGS, biological and ecological observations of the DOI, synoptic meteorological and bathythermograph data from the DOD, fossil fuel statistics from the DOE, demographic data from the Bureau of the Census, and soil maps from the USDA Soil Conservation Service. Table A2 provides examples of data sets that are needed to help answer specific global change research questions.

From the perspective of such other programs and projects, global change research constitutes a secondary use whose needs are generally not funded. Only some of the crucial programs and projects have already been identified as USGCRP *contributing* programs; their budgets in FY 1993 were about \$1.4 billion—about equal in size to the total for the focused USGCRP.

**Table A2. Examples of Data Needed from Outside the USGCRP**

Topical Questions*	Data and Information Examples
How will climate change affect temperature, precipitation, soil moisture patterns, and the general distribution of water on the land surface, and how will these changes in turn influence the atmosphere?	<ul style="list-style-type: none"> <li>• Streamflow</li> <li>• Soil moisture</li> <li>• Ground and surface water</li> <li>• Precipitation</li> <li>• Evapotranspiration</li> <li>• Water use</li> <li>• Temperature</li> </ul>
How do different coastal regions respond geologically and ecologically to rising sea level, and how can the contributions from changes in climate be differentiated from those due to tectonic processes?	<ul style="list-style-type: none"> <li>• Ground level motion</li> <li>• Height of waves on shores</li> <li>• Sea level</li> <li>• Sea and land ice thickness</li> <li>• Sediment cores</li> </ul>
How do the oceans interact with the atmosphere in the storage, transport, and uptake of heat?	<ul style="list-style-type: none"> <li>• Buoy data</li> <li>• Ocean ship histories</li> <li>• Ocean ship transect data</li> <li>• NOAA and DOD satellite data</li> </ul>
In which ecological systems and species are significant changes most likely to occur, and what attributes of importance to humans will be at risk?	<ul style="list-style-type: none"> <li>• Migration and productivity</li> <li>• Long-Term Ecological Research (LTER) data</li> <li>• Soil classification</li> <li>• Ecotone maps</li> <li>• Status of threatened species</li> </ul>
What are the natural ranges and rates of change in the climate and environmental systems?	<ul style="list-style-type: none"> <li>• Temperature</li> <li>• Ice cores</li> <li>• Cloud records</li> <li>• Precipitation</li> <li>• Ground-based solar grid</li> <li>• Tree rings</li> <li>• Upper air data</li> </ul>
What are the effects of volcanic emissions on regional and global climate?	<ul style="list-style-type: none"> <li>• Volcanic gases and aerosol emissions</li> <li>• Radiative forcing</li> </ul>
What is needed to develop the information required to verify models of interaction between human and natural systems and to assess the likelihood of changes in those processes?	<ul style="list-style-type: none"> <li>• Demographic and census data</li> <li>• Health statistics</li> <li>• Land use</li> <li>• Economic activity and trade</li> <li>• Mineral use</li> <li>• Technological change and distribution</li> <li>• Pollutant distribution and effects</li> <li>• Emission sources and sinks</li> <li>• Energy production and use</li> </ul>

\* Questions taken from *Our Changing Planet: The FY 1991 Research Plan*.

From Table A2 it can be seen that not only are the sources of such data and information from many agencies, but also that they cover all the central priorities of the USGCRP. Such data and information are critical not only in their own right for global change research, modeling, and assessments, but also to the focused data and information program to fill gaps in coverage, tie together diverse data sets, and improve the quality and usefulness of the data and information by providing ground truth at selected points for calibration. The latter is particularly important for the remotely sensed data required by the USGCRP. Also not apparent is that some of these critically important data and information are unavailable to the USGCRP because they are intermingled with material subject to security, proprietary, or regulatory constraints.

This special issue identifies the need for agencies participating in the USGCRP to assemble, document, archive, and disseminate the data and information critical to the USGCRP produced by programs that do not have global change research as a primary objective. The goal is to provide GCDIS access in useful form to data from such programs and projects through existing and planned data systems and centers or by creating new data centers for them. Within the framework established, agencies will propose individual or interagency activities to meet specific USGCRP requirements. These requirements then will be coordinated from an interagency perspective through the CENR. By using this coordinated approach, the program will reap the benefits of using data and information from outside the USGCRP and allow for deeper insights that arise from cross-disciplinary analysis.

Although the existing USGCRP priorities and milestones help focus the data capture activities, many thousands of related data and information products of potential relevance still exist, with many needing an investment of resources to become well documented and accessible. As a result, it is essential to establish more specific priorities for data capture activities. This will be done in cooperation with the other elements of the USGCRP, the user community, and advisory groups such as the NAS.

Agency planning for the special issue recognizes that global change data management efforts should be well coordinated with other established interagency data management activities, such as the FGDC, the Office of the Federal Coordinator for Meteorology, and other appropriate coordinating organizations and committees.

## Data Capture Activities

To make relevant data and information resulting from data capture activities available to the USGCRP, a number of different activities are required. These are grouped here into the four activity areas described in *The U.S. Global Change Data and Information Management Program Plan*: that is, *assembling, documenting, archiving, and disseminating*. Specific activities that overlap activity areas are included in the area that is expected to require the most resources.

### **Assembling**

1. *Inventory of data.* Agencies need to set specific objectives and conduct an inventory of previously generated data and information that focuses on those research areas most relevant to the highest priority science objectives and research milestones of the USGCRP. Library holdings will be included in the inventory. (It is estimated that more than half the publications in Federal scientific and technical libraries and information centers relate to global change.) Attention will be given to obtaining input from the NRC and other representatives of the user community in order to guide development of the inventory.
2. *Inclusion in the GCDIS.* Data and information included in the GCDIS will be based on the USGCRP priorities and the advice of the user community. Where agencies cannot provide direct access to their holdings, funds will be required for data transfer to data centers participating in the GCDIS. In some cases, extracts will be created, eliminating material that is sensitive due to security, proprietary, regulatory, or other considerations. In other cases, key data sets will be put into digital form in order to be generally useful.
3. *Integrate preexisting programs.* Many of the focused USGCRP programs were started before the widespread appreciation of the need to plan for interagency data and information management in the programs. For example, resources are needed to integrate with the GCDIS the data and information management components of programs such as the TOGA, the WOCE, and the GEWEX.
4. *Data rescue.* Many data and information products of vital importance for global change research need immediate rescue and long-term maintenance to avoid deterioration and loss.

### **Documenting**

5. *Priorities.* It will be almost impossible to document adequately all data and information potentially useful for global change research so that they can be properly used decades after their creation. With participation of the NRC and the user community, priorities will be established for the data and information

requiring such documentation—taking advantage of available expertise before such opportunities disappear forever.

6. *Current data center holdings.* The adequate documentation of holdings already in data centers is a major undertaking. The most critical holdings need sufficient documentation so that they not only will be accessible, but can also be used with confidence decades from now. This level of documentation is typically much more stringent than now exists and requires significant effort to remedy.
7. *Data from other sources.* Global change studies often require correlating data from disparate sources, such as remotely sensed data combined with in situ measurements. Even within a single set of data there are often discontinuities attributable to instrumentation changes or other events. It is essential that complete information describing these data is available to researchers in order that the data can be applied appropriately.
8. *Training.* Training of agency research staffs, information specialists, and managers will be required to institute the policies and procedures for managing data and information to support global change research, developing the skills needed for effectively using the wide range of GCDIS capabilities, and improving communication of technical specialists with the public. Attention will be given to creating global change research network information, information products, and electronic data base information appropriate to educators at the K–12 levels.

## Archiving

9. *Long-term storage facilities.* Agencies participating in the GCDIS have agreed to manage all data and information in a manner that complies with archiving standards. Agency data centers will need significant additional resources to handle not only the large numbers of additional data and information products that need to be archived, but also—for such programs as the Next Generation Weather Radar (NEXRAD)—very large data volumes.
10. *Small data centers.* Libraries and small data and information centers will collect more and more data sets and information created through individual research efforts. These data and information will sometimes be available at Federal libraries, but more often they will be found at academic institutions. Such local reference and research systems need to be compatible with the GCDIS, and many need be upgraded in order to store, retrieve, and send large amounts of data and information to users.
11. *Interagency coordination.* The GCDIS will have a wide range of interagency coordination requirements, including inventory search and order, pricing, order tracking and billing, user services, submission guidelines, system

interfaces, retention and purging, media, formats, and performance assessments. As the amount and scope of the data and information in the GCDIS increases, support for such interagency and external coordination will increase.

12. *Researcher feedback.* Agencies will actively involve users and others in the affected research communities in their global change data and information management activities. This will include periodic formal reviews by expert panels such as the NAS CGED, as well as surveys of user satisfaction and the soliciting of suggestions for improvement of the GCDIS.

#### **Disseminating**

13. *User services.* Data and information centers and libraries will need resources to provide the data and information to users and to respond to user questions both on the proper use of specific data and on the holdings of other data centers. This need will increase with the number of products in the GCDIS and with the number of its users. Among the Federal and academic libraries, as well as the Federal Depository Libraries, access to the Internet and CD-ROM equipment is essential.
14. *Consolidated services.* In some cases, data center services for data capture data and information exist but are so fragmented and incompatible with the GCDIS that they are almost unusable. For example, there are 18 separate Long-Term Ecological Research Centers that can only be accessed individually. The GCDIS will not only tie all the participating data centers together, but also create dissemination mechanisms to allow researchers to deal with groups of centers as a single entity.
15. *International.* Access to data and information holdings outside the United States that are critical to the USGCRP will require significant resources. For instance, data and information sets from international sources may need to be obtained by connecting them to the GCDIS through international mechanisms such as the UNEP Global Resources Information Database (GRID). Additionally, selected technical information received from international sources needs translation into English.
16. *Standards.* Compliance with national and international standards in use by libraries and data centers is crucial even though it may be costly. Access to major library collections, information centers, and bibliographic information systems is dependent upon implementation of standards developed for bibliographic control, communications, product formats, and product delivery.

## Fiscal Requirements

Although the annual USGCRP program plans have recognized a dependence on data and information from programs not focused on global change, resource requirements have not previously been identified to assemble, document, archive, and disseminate these crucial data and information. Since these data and information reside with groups whose primary mission is not global change research, in many cases an additional specific allocation of resources is needed for these data and information to be made available to support global change research.

From the perspective of the agencies, the funding required by the special issue on data capture is over and above that which the agencies had previously planned as their part of the USGCRP. The individual IWGDMGC agencies, however, have been encouraged to submit budget requests to meet this critical need. Agency special issue plans will be a part of each agency's GCDIS implementation plan.

## **B. Agency Implementation Overviews**

### **The Department of Commerce**

#### **Summary**

NOAA is the primary agency within the DOC for GCDIS participation as both supplier and user of climate and global change data and information products. Other DOC agencies will participate in the GCDIS as their roles are better defined and as the system evolves. For example, the U.S. Bureau of the Census, through its Center for International Research, has initiated global change, data-related, collaborative activities with the CIESIN. The Census Bureau is also attempting to increase public awareness of and use of its own data through the development of its own Internet site and other online electronic collaboration with appropriate parties.

NOAA's three NDCs, and the NOAA Satellite Active Archive, the NOAA Directory Services, the NOAA Library, and the NOAA NIC are agency components that will participate in GCDIS operations. As the GCDIS evolves, NOAA, through its NOAA-wide Data System Modernization, will add additional capabilities and improved interoperability for the GCDIS user community. NOAA's Data System Modernization is now in the planning stage; system implementation is expected to start in 1995 and continue through at least 2004.

#### **NOAA GCDIS Capabilities Available April 1, 1994**

Serving as the Nation's steward for the long-term climate data record, NOAA will provide GCDIS support through its NDCs—the National Climatic Data Center, the National Geophysical Data Center, the National Oceanographic Data Center—and their associated Solar, Geophysical, and Environmental WDCs. Each center provides a customer service facility and online access to metadata and data sets. Online access to operational satellite data is provided by the NOAA Satellite Active Archive. User access to NOAA-wide data and information is also available through the NOAA Directory Services and the NIC. The NOAA Environmental Services Data Directory provides users with online access to descriptions of more than 2,500 NOAA data sets. The NOAA NIC provides Internet access for users seeking information about NOAA and its data and information holdings. The NOAA Library provides GCDIS users with online catalog access to 1.5 million volumes, including 10,000 volumes of historic worldwide meteorological and global change climatological data.

**NOAA GCDIS Components Available by April 1, 1994**

*National Climatic Data Center (NCDC) data content.* The NCDC receives data from national operational observing systems managed by the National Environmental Satellite, Data, and Information Service (NESDIS), NOAA, the military services, the Federal Aviation Administration, and the U.S. Coast Guard. Foreign meteorological and climate data are received from international sources through the Global Telecommunications System of the WMO's World Weather Watch Program and through exchange of foreign data through the WDC system. Designated as the WDC-A for Meteorology, the NCDC also collects and manages the data bases of international experiments such as the WCRP sponsored by the WMO and the ICSU. In addition to the observational data, the NCDC maintains an archive of modeled meteorological data and products produced by NOAA centers and laboratories and by international experiments.

Major in situ data sets archived at the NCDC include U.S. hourly surface weather observations (temperature, pressure, precipitation, weather, wind, etc.), global 3- and 6-hourly surface weather observations (land and marine surface ships), U.S. and global upper-air observations (temperature, wind, humidity, and height for various pressure levels), daily and monthly data (maximum, minimum, and mean temperature and precipitation) for approximately 7,000 U.S. locations, NOAA atmospheric trace gas data, weather radar photographs and other digital data from the new WSR88D national radar system (NEXRAD), and published daily and monthly data for foreign locations. Most of the data since 1949 are in digital form and (from before 1949) on hard-copy media.

Satellite data are archived by the NCDC and its Satellite Data Services Division (SDSD). The archive consists of full-resolution and derived products from NOAA operational polar and geostationary satellites dating back to 1978. Non-NOAA satellite data are also archived, such as DMSP special sensor microwave data. A full-resolution Geostationary Operational Environmental Satellite (GOES) data archive is managed through a contract with the University of Wisconsin. Both the SDSD and the University of Wisconsin have customer service groups.

NOAA is improving access to satellite data through the Satellite Active Archive (SAA). Internet access, browse, search and order, and Internet data set file transfer capabilities are available for selected AVHRR data sets. The SAA is also interoperable with NASA's Version 0 EOSDIS. Internet access is also available for NOAA-generated Pathfinder Data Sets. The University of Wisconsin also has GOES browse images and selected NOAA-NASA Pathfinder Data Sets online, accessible through a Gopher server.

The NCDC is preparing Baseline Data Sets to support research and monitoring of long-term climate trends. Some of these data sets are the Global Baseline Data Base of Atmospheric Trace Constituents, the Global Historical Climate Network (GHCN), the Comprehensive Aerological Research Data Set (CARDS), and the Comprehensive Ocean Atmosphere Data Set (COADS). Data are being prepared in a common format and processed using standard quality control processes.

A Global Climate Perspectives System (GCPS) is being developed and implemented at a limited number of research locations. The GCPS will house several global climate data sets and provide a data base management system, analysis software and graphical software packages, and data export capabilities. Researchers will be able to access the system through the Internet; have access to data sets; and be able to perform analyses, display analyses and long-term trends, compare data parameters, and export and download data across the Internet. The capabilities of this system will allow researchers to put current climate trends into perspective relative to the entire climate history.

The NCDC will also operate a rudimentary NOAA Environmental Watch (NEW) project. The NEW is aimed at integrating a variety of environmental data sets to produce primarily high-level, policy-oriented information products on a variety of media, including CD-ROMs, diskettes, hard copies, and electronic access.

*NCDC access.* Data are available by telephone, mail, automated facsimile, World Wide Web, WAIS, and online through the NCDC Online Access and Service Information System and the GCPS. Data sets available online are limited to most recent 2 years of data for U.S hourly surface observations, upper-air observations, NEXRAD data inventories, NOAA Geophysical Fluid Dynamics Laboratory modeled data sets, NOAA Global Climate Model data sets (100-year output), subsets of the NOAA Atmospheric Trace Gases data set, and daily data from 7,000 U.S. locations. Online access capabilities include limited browse capabilities, online ordering, and online and offline data delivery. Baseline Data Sets, the COADS Release 1a, the CARDS subset (kernel data set) and the GHCN Version 1 are available on magnetic media. A Research Customer Service Group provides customer services for research users requiring additional assistance in acquiring data or information about NCDC data sets.

*National Geophysical Data Center content.* The NGDC, in conjunction with its collocated WDCs (Paleoclimatology, Solid Earth Geophysics, Solar-terrestrial Physics, Marine Geology and Geophysics, and Glaciology–Snow and Ice), distributes a wide variety of global change data using many types of media and systems. NGDC's Paleoclimate Program distributes pollen count data, tree ring measurements, ice core accumulations, and fossil plankton counts, as well as historical records. The NGDC archives and disseminates data from the Defense Meteorological Satellite

Program (including visible, infrared, and special sensor data) for global change investigations. The NGDC also distributes a suite of global environmental data bases including the Global Ecosystems series of CD-ROMs. These global data sets include land use, wetlands, vegetation, climate, global relief, soils, and outputs from various global environmental models. The NGDC, through the National Snow and Ice Data Center at the University of Colorado, provides global change data relating to the cryosphere. It disseminates such data as snow cover, sea ice, glacier mass balance, and information describing polar climatology. Other relevant global change disciplines at the NGDC are solar variability and the sea floor environment. Specific examples of data sets available from NGDC are lake and marine sediment, ice core, and tree ring data; topographic and bathymetric measurements, data illustrating natural hazards such as earthquakes, tsunamis and volcanoes; and historical sunspot activity data.

*NGDC access.* All NGDC data and derived products are distributed online, on various media such as .5-in. and 8-mm magnetic tapes, floppy diskettes, CD-ROMs, or in hard-copy forms such as microfilm, paper, or 35-mm slides. Access to NGDC data is by mail, telephones, electronic mail, fax, personal contact, and through online activities. NGDC's online services include descriptions of NGDC data and services, phone lists, selected data sets, and sample images of data products. NGDC's online access system, using current technology, such as the Internet, Gopher, World Wide Web, and others, list many of NGDC's current data holdings, both as directory and guide information. A number of NGDC data bases are also available through the Internet, and these include all of NOAA's paleoclimate data as well as selected solar variability, global ecosystems, and marine environment data.

*National Oceanographic Data Center (NODC) content.* The NODC receives observations, publications, and other data and information products from nearly every country involved in oceanographic research. Through a system of bilateral data exchange agreements, and under the auspices of the WDC system, observational data and data products are routinely shared with more than 30 countries. Under separate agreements, the NODC also manages data for several multidisciplinary experiments operated under the WCRP.

Data received, processed, and archived include in situ observations, remote sensing data products, and research reports and publications. Data received cover the breadth of ocean science, ranging from bathythermograph and Nansen casts to subtle ocean chemistries and plankton counts; and span the global ocean from the late 1800s to the present. The NODC archives are the largest single collection of global oceanographic data covering the longest time span. Satellite and other remote-sensing data types (such as acoustic Doppler current profiles) have been added to NODC archives during the past few years and, with adequate funding support, should be available to the GCDIS.

With the support of the NOAA Climate and Global Change program, the NODC is preparing new ocean climatologies on a geographical and temporal scale never before available. The major physical oceanography data bases have been merged, quality-controlled, and statistically treated to produce a global ocean climatology for one-degree squares. In addition to the climatologies, a set of long time series observations for specific areas in the ocean and the largest and most frequently used data bases are being published on CD-ROMs, and are available online in limited quantities through the Internet and through dial-in access.

*NODC access.* Access to the entire holdings of the NODC are available by telephone, mail, Internet order, and fax. Funding permitting, online access is restricted to physical oceanography data bases including bathythermographs, hydrocasts, current meter, time series hydrocasts, Conductivity Temperature Depth/Salinity Temperature Depth, and one or more major experimental data bases (e.g., WOCE, JGOFS). In addition, a comprehensive catalog of holdings and the NOAA-USGS-EPA Taxonomic Code will be available.

*NOAA Directory Services.* NOAA Directory Services include four data bases. The NOAA Environmental Services Data Directory (NOAADIR) has more than 2,500 descriptions of NOAA information and data sets in Directory Interchange Format (DIF). The data sets described contain information about the atmosphere, oceans, climate, hydrology, geology, geophysics, and solar terrestrial physics. There are also data sets concerning astronomy, biology, diseases, and health issues. Most of the data sets regarding biology come from the NOAA National Marine Fisheries Service (NMFS) and the Foreign Meteorological Collection in the NOAA Library. The library collection contains records of weather observations and geologic phenomena, as well as records that enhance the human dimensions data available for global change studies. Some of these records cover demography, diseases, health issues, morbidity, mortality, and epidemics. So far, more than 800 data records in the library have been documented in the DIF format, with about 2,000 remaining to be done.

Other data bases that are available in the NOAA Directory Services are the NOAA National Environmental Data Referral Service (NEDRES), the NOAA Product Information Catalog (PIC), and documents for the NOAA National Environmental Watch.

The NOAA Environmental Services Data Directory will be available through two versions of the Master Directory (MD) software (as supplied by NASA), as well as through the WAIS. The NEDRES, the PIC, and the NOAA National Environmental Watch will be available through the WAIS. Users can access any of these data bases through dial-in services, including local phone service and a 1-800 service, or through the Internet. Access to the MD software and the WAIS will also be provided through a Gopher server.

**NOAA GCDIS Components Added by April 1, 1995.**

All NOAA components participating in the GCDIS will add incremental enhancements to data content and access functions for the GCDIS user community. A major area of enhancement will be improving online access to NOAA data. Summaries of some of these enhancements follow.

The NCDC will add data to expand the period of record for online data sets, geographical coverage will be expanded, and high-interest data sets will be added. Browse and visualization capabilities will be enhanced. Information products will be available on the Internet through the NEW project. NGDC's Defense Meteorological Satellite Program (DMSP) digital imagery data will be available online for browse, visualization, access, and online and offline ordering. NGDC's paleoclimatology data base will have its regional spatial and temporal resolution increased while greatly improving online accessibility. The NODC will add chemical oceanography, biologic data containing primary productivity, and chlorophyll data sets. New releases and versions of the Baseline Data Sets will be added for online access at the NCDC. The NOAA SAA will expand the number and time period of operational and Pathfinder Data Sets available for online access. More than 6,000 DIFs will be added to the NOAADIR. Many of these DIFs will come from the Foreign Meteorological Collections at the NOAA Library, the NCDC, and the NMFS. The DIFs from the NOAA library will include many human dimensions parameters, whereas the NMFS's includes many biological parameters.

**NOAA GCDIS Components Added by April 1, 1996.**

Baseline Data Sets, the COADS Release 2, the CARDS Version 1, and the GHCN Version 2 will be available at the NCDC. The GCPS will be implemented and accessible to an expanded number of research groups. The NEW project will have its information content about environmental issues expanded to include major scientific assessment information, such as IPCC Climate Change and WMO/UNEP Ozone Assessments. Major enhancements will be made to high-priority global change data sets at the NGDC, such as paleoclimatology, DMSP, solar variability, global snow cover, and global ecosystems. The NODC will add acoustic Doppler current profiler data. The NOAA Directory Services will add another 3,000 DIFs, and will be linked to the EOSDIS, with interoperability through the EOSDIS and the NOAA Satellite Active Archive.

**NOAA GCDIS Functionality: Access Capability Table**

This table shows the integration of NOAA's data and information access capabilities for 3 years. Some of NOAA's GCDIS components will be operating at higher levels, and some at lower levels.

NOAA	1994	1995	1996
Archive	2	2	1
Directory	2	2	2
Guide	4	3	3
Inventory	4	3	2
Browse	4	3	2
Order Place	2	2	2
Distribution	3	2	2

## Department of Defense

### DOD GCDIS Capabilities Available April 1, 1994

The DOD routinely collects environmental data globally in support of DOD operations. The DOD also conducts mission-related research into environmental processes and conditions that affect defense operations, tactics, and systems. The DOD does not have a mission requirement to archive data for civil uses. DOD data products are made available to the appropriate U.S. national archives for subsequent use by the GCDIS.

### DOD GCDIS Components Available by April 1, 1994

*GCDIS content: holdings available and identification of major data sets.* The Commander, Naval Meteorology and Oceanography Command, supports activities that provide meteorological, oceanographic, and mapping, charting, and geodetic products for DOD and U.S. Navy operations. Complete synoptic and aviation meteorological observations, bathythermograph and STD casts, bathymetry, and limited biological observations are routinely recorded. Routine synoptic meteorological measurements and bathythermograph data, plus a full set of air-ocean analysis and forecast products, are produced by the Fleet Numerical Meteorology and Oceanography Center. Special Sensor Microwave Imager (SSM/I) Sensor Data Records, Temperature Data Records, and Environmental Data Records on an orbital basis for the DMSP are generated by this center.

The National Ice Center, a cooperative effort among the Navy, NOAA, and the Coast Guard, provides routine ice products to military and civil users and provides global ice observations, analyses, predictions, and advisory information to military and civilian weather forecast offices with sea/lake ice responsibilities. Data parameters include ice edge, concentration, age, and temperature, as well as features such as ridges, leads, and polynyas.

Operating Location-A (OL-A), United States Air Force Environmental Technical Applications Center, collocated with the NCDC, archives observational and analysis data for military purposes. Observational data sets archived at OL-A include Foreign Synoptic Data, Airways/METAR (Airport Weather) data, Foreign PIBAL (Pilot Balloon) data, and Worldwide Radiosonde and Rawinsonde data. Analysis data sets available at OL-A include Real Time Nephanalysis (RTNEPH) and 3DNEPH (precursor to the RTNEPH), the Air Force Global Weather Central (AFGWC) High Resolution Analysis System (HIRAS) data, and AFGWC surface temperature data sets. The HIRAS includes worldwide analysis of pressure, wind, temperature, moisture, and other analysis variables at mandatory pressure levels. The USAF DMSP satellites continually collect visual and infrared environmental data. The USAF and NASA initiated the effort to capture digital DMSP data at AFGWC on magnetic tape for archiving at the NGDC. Implementation will be in March 1994.

The USAF also sponsors the Cloud Information Reference Library and Archive (CIRLA), a reference archive containing information on cloud and cloud-related environmental data. CIRLA was designed in response to the needs of DOD systems development and operational communities, but CIRLA is also responding to increasing emphasis on requirements to transfer environmental data from the DOD to the civilian scientific and engineering communities.

The U.S. Army Corps of Engineers maintains the Cold Regions Research Engineering Laboratory, which conducts research on snow, ice, and frozen ground and the effects on construction, transportation, and military operations. Additional data collected by the Department of the Army include meteorological observations collected by Army Meteorological Teams and observations made by the Army Corps of Engineers in support of various civil works projects. Complete synoptic-form meteorological observations, together with soil temperature, solar radiation, and optical transmissivity, are recorded. Temperature, salinity, water quality, precipitation, river stage, wave height, wave period, wave direction, and beach erosion rates are also routinely collected by the Army Corps of Engineers in support of specific continuing projects.

The Defense Mapping Agency (DMA) is the DOD manager for mapping, charting, and geodesy (MC&G) data around the world, and adopts or establishes MC&G standards promoting interoperability within the DOD and with allies. In response to

DOD requirements, the DMA compiles, produces, reproduces, and distributes maps, charts, and digital MC&G data. These types of data aid in determining watershed and shoreline changes, sea level rise, variations in coastal currents, rates of urbanization, and changes in the extent of vegetated areas. Release of DMA data depends on security classification, level of resolution, and any third-party restrictions that may be in effect.

The Defense Technical Information Center (DTIC) is the central repository within the DOD for scientific and technical information generated in support of DOD research and development efforts. The DTIC primarily serves the DOD, other Federal government agencies, and their contractors. The DTIC forwards to the National Technical Information Service (NTIS) all DOD technical information provided to the DTIC by DOD organizations when cleared for public release. The NTIS serves the general public by providing access to these data for DOD and the other Federal research and development agencies.

Research organizations including the Naval Research Laboratory, NASA, the Jet Propulsion Laboratory, the NSF, Rand Corporation, the Army Engineering and Topographic Lab, the Air Force Institute of Technology, and Rice University provide global change information for the DTIC to disseminate. Information on global change in DTIC's collection includes analytical reports about research conducted, as well as data sets, models, projections, research proposals, and summaries of research efforts. Specific research efforts are varied and include such topics as the impact of global climate change on the U.S. Navy, sea level and sea ice changes, a CD-ROM data base containing worldwide meteorological data, global climate modeling, temperature and trace gas trends, and global cloud analysis algorithms.

Basic metadata, described as level 3 in the implementation plan, are currently provided by the DOD. Level 1 or 2 metadata are beyond DOD's mission. The DOD may, however, on a case-by-case basis, consider providing metadata at a higher level if a DOD public domain data set is identified to be an extremely high priority for the study of global change.

*GCDIS access.* Access to DOD unclassified and unrestricted data and information is provided through national archive centers: the NCDC, the NODC, the NGDC, the National Snow and Ice Data Center, and the NTIS. In addition, the USGS and NOAA serve as distribution agents for DMA products releasable to the public.

The DOD and NOAA share responsibilities for processing polar-orbiter satellite data and exchanging unclassified satellite data and products using the Shared Processing Network (SPN). There are four centers of expertise that process satellite data: the Fleet Numerical Meteorology and Oceanography Center for microwave products, the NAVOCEANO for sea surface measurements, the AFGWC for visible and infrared

cloud products, and NOAA's NESDIS for atmospheric sounding data. The NESDIS is also responsible for archiving any SPN data it receives and for distributing those data and products to civilian users.

#### **DOD GCDIS Components Added by April 1, 1995**

The DOD will not add GCDIS components.

#### **DOD GCDIS Components Available by April 1, 1996**

The DOD will not add GCDIS components.

#### **DOD GCDIS Functionality: Access Capability Table**

The agencies that receive DOD data products, primarily NOAA and the USGS, assume the responsibility for GCDIS functions, such as archive, directory, distribution, networks, user support, and others.

## **Department of Energy**

### **Summary**

The DOE's contribution to the GCDIS comprises three focused programs: the ARM Archive, the CDIAC, and the Program for Climate Model Diagnosis and Intercomparison (PCMDI). The ARM archive features data from measurements and observations of radiative and atmospheric phenomena, as well as the ancillary information and documentation that allow scientists to interpret the data. The CDIAC provides data and other information concerning carbon dioxide, other greenhouse gases, and climate change to the research, policy-making, education, and corporate communities, and to the public. The PCMDI archives a broad range of global data sets of gridded meteorological data to serve in the validation of climate models. The CDIAC and the ARM Archive data and information products (all provided free of charge) are presently provided in response to mail, telephone, fax, or email request. These data are available through the Internet by ftp or on a variety of physical media to meet the needs of the users: 8-mm tape, nine-track tape, or floppy disk.

#### **DOE GCDIS Capabilities Available April 1, 1994**

*The ARM Archive content.* The ARM program is a DOE-sponsored global change research effort designed to improve the modeling of cloud radiative forcing in general circulation models. The primary user community for ARM data is the atmospheric research community; secondary users are the broader scientific

community that has interest in some of the meteorological or radiative measurements that the ARM can provide.

The ARM Archive receives and manages a broad variety of data, such as

- Measurements and observations of radiative and atmospheric phenomena,
- Information and documentation about the data streams that allow scientists to interpret the data,
- Information about the instruments producing the data, and
- General information about the ARM project.

This information comes from a variety of sources within the ARM project and from several external sources. The primary data source is the ARM CART, which will consist of three highly instrumented sites located worldwide. The first site, centered close to Lamont, Oklahoma, has been taking data since June 1992. Data fusion products and data quality measurement products are generated at the ARM experiment center using data from the CART sites and observations from such external sources as satellites. Information about instruments, data quality, and instrument operations comes from instrument developers, instrument mentors, site operations staff, and scientists.

*ARM Archive access.* The ARM Archive data and information products (all provided free of charge) are presently provided in response to mail, telephone, fax, or email request. The data are available through the Internet by ftp or on a variety of physical media to meet the needs of the users: 8-mm tape, nine-track tape, or floppy disk.

The ARM Archive will continue to enhance the number and types of scientific data available as the ARM project continues to add instruments and sites. In addition, the ARM Archive will increase and enhance the quality and quality of information available to assist in the understanding and interpretation of the ARM data.

An automated user interface will be available early in 1994 for requesting ARM data and information files. This interface will allow users with X-windows-based workstations to specify the location, instrument platform, date range, and data processing level for the data they want to receive. Once that request is submitted, the ARM Archive Mass Storage System will retrieve the requested files and notify the user of their availability for transfer by ftp. If the user requests different delivery media, the system will package the data onto those media and notify an operator to ship the media to the user. Finally, the ARM Archive will maintain general information about the ARM project, instruments, and CART sites on a World Wide Web information server.

*Carbon Dioxide Information Analysis Center data content.* The CDIAC provides data and other information about CO<sub>2</sub>, other greenhouse gases, and climate change to the research, policy-making, education, and corporate communities and to the public. Thus, CDIAC's mission is broad, and the suite of data and information products provided is correspondingly broad. CDIAC's data holdings, for example, include concentrations of greenhouse gases in the atmosphere and oceans, carbon storage in ecosystems, greenhouse gas emissions from industry and land use change, long-term global and regional climate, response of vegetation to elevated CO<sub>2</sub>, tree ring data, and data on the vulnerability of coastal areas to rising sea level.

The CDIAC acquires data from a large number of investigators and institutions in the United States (not limited to DOE-funded programs) and around the world. Internationally, the CDIAC obtains climate and greenhouse gas data, in particular from China and Russia, through its participation in bilateral research agreements. The CDIAC also operates the WDC-A for Atmospheric Trace Gases, a component of the WDC system coordinated by the ICSU.

The CDIAC quality assures, documents, archives, and distributes data in the form of numeric data packages (of which approximately 50 currently exist). When a particular data base does not exist, the CDIAC compiles the data from various sources (e.g., its global and country-by-country data base on emissions of CO<sub>2</sub> from fossil fuel use and cement manufacturing).

In addition to numeric data packages, CDIAC products include the newsletter *CDIAC Communications*, the series *Trends: A Compendium of Data on Global Change*, the *DOE Research Summary* series, the *Glossary: Carbon Dioxide and Climate*, and specialized bibliographies. The CDIAC also distributes global-change-related reports from the DOE and other agencies.

*CDIAC access.* The CDIAC data and information products (all provided free of charge) are currently provided in response to mail, telephone, fax, and email requests. In addition, direct access to CDIAC's data and documentation files is now available electronically in its anonymous ftp area. The CDIAC data are provided on a variety of physical media to suit the needs of the users: printed reports and hard-copy listings, 9-track magnetic tapes, floppy diskettes, 8-mm tapes, and CD-ROMs.

#### **DOE Components Added by April 1995**

In late 1994 or early 1995 the ARM Archive will provide online, searchable, textual information about instrument operation and data quality. Also, information bearing directly on a set of requested data will be packaged and delivered with the data. Both of these capabilities will be available to users through the automated user interface.

The CDIAC will expand the breadth and depth of its data holdings, particularly in the areas of greenhouse gas concentrations in the oceans and atmosphere and emissions of greenhouse gases from land use changes.

The CDIAC expects to benefit from its collocation with the NASA EOSDIS DAAC for Biogeochemistry, also located at the Oak Ridge National Laboratory (ORNL). As the NASA EOSDIS matures, it is expected that hardware and software systems developed by and for the ORNL DAAC will also provide more online access to CDIAC data products.

#### **DOE Components Added by April 1, 1996**

The CDIAC will continue to expand its data holdings, responding to the changing needs of the global change community. As the GCDIS approaches greater operability, information system components such as the GCMD will provide enhanced online access to CDIAC data products. Further, it is possible that the CIESIN will serve as an additional conduit providing access to CDIAC data products, especially for the social sciences and the international communities.

DOE	1994	1995	1996
Archive			
Directory			
Guide			
Inventory			
Browse			
Order Place			
Distribution			

## Department of the Interior

### Summary

As the major Federal land manager and primary Federal agency responsible for managing the Nation's natural ecosystems, fish and wildlife, and energy and water resources, the DOI is particularly concerned about the potential short- and long-term effects of climate and other environmental change on these lands and resources. The DOI's global change research is addressing topics such as hydrologic and geologic processes and resources, land use, land cover, biological habitats, resources, and diversity; past global change recorded in the physical, chemical, and biological record; land surface and solid-Earth processes that relate to environmental change; geography and cartography; polar and arid region processes; ecosystem modeling and dynamics; and resource ethnology.

The DOI bureaus collect, maintain, analyze, and interpret short- and long-term land, water, air, biological, and other natural resource data and information in support of their missions. These efforts have always included maintenance of high-quality, long-term data sets, including cartographic, land cover, geologic, hydrologic, ecological, and biological data from both satellite- and aircraft-based remote-sensing and terrestrial-based observations. The DOI will provide access to directory-level descriptions of these data (and inventory-level information where feasible) through DOI and other agency directories and clearinghouse mechanisms, such as the NOAA National Geophysical Data Center. The USGS GLIS will be one of the primary inventory-level interfaces with the GCDIS and an access point to DOI global change data. The DOI also participates in the NASA EOS program through the EOSDIS Land Processes DAAC at the USGS EROS EDC, where capabilities are being developed to process, archive, and provide online information system access to EOS land-related data sets such as those from the MODIS.

### DOI GCDIS Components Available by April 1, 1994

#### GCDIS Content

*The USGS EROS Data Center.* The USGS's EROS EDC is a core element of DOI's focused global change data management activities. The EDC has archived, processed, and distributed Landsat data since the launch of Landsat-1 in 1972. With the enactment of the Land Remote Sensing Policy Act of 1992, the Center was designated as the National Satellite Land Remote Sensing Data Archive. This designation has broadened the Center's responsibility for preserving and providing long-term access to satellite-acquired remotely sensed data of the Earth's land areas. The Center's 21-year collection of approximately one million Landsat scenes is the largest component of the archive. These data are currently being converted from

aging magnetic media to new cassette tapes to preserve them for future use. Current activities include conversion of approximately 300,000 scenes of post-1978 MSS data (30 percent of these data were converted by the end of 1993) and 170,000 scenes of Landsat Thematic Mapper data, and the development of improved Landsat data processing capabilities.

Other types of satellite data are being added to the EROS Data Center's Archive. Since 1992, the USGS has been working with NASA, NOAA, the ESA, and more than 20 foreign ground receiving stations to collect 1-km resolution AVHRR data for each daily pass over the Earth's land surface. These data support many applications, including efforts by the global change research community to produce a global land cover maps and to monitor vegetation conditions (greenness) on a periodic basis throughout the year. This program is being coordinated with the IGBP and the CEOS. It complements the operational NOAA programs that collect global AVHRR data at 4- and 16-km resolution, primarily for oceanic and atmospheric applications.

Long-term, consistent continental and global land data sets are being developed and produced to meet the needs of global change researchers. Emphasis is being given to data sets related to land use and land cover, vegetation, soils, terrain, and satellite-derived data, such as global vegetation index and albedo (brightness) data. The application of these data sets to process research is being tested in several projects, including modeling of land-atmosphere exchange processes, analysis of vulnerability of arid and semiarid land to changes in long-term productivity, and biogeochemical exchange processes in Alaska and the Arctic.

*USGS Global Change and Climate History Program Data Management System.* The USGS is providing Internet access to data and information resulting from its Global Change and Climate History (GCH) Program through an online system. Users can obtain information about the GCH Program and its projects, data sets resulting from GCH-funded research, data from other sources of interest to USGS global change researchers, and freely distributable software tools that may be used to obtain and interpret data. The GCH Data Management System represents an effective approach for making the evolving results of GCH research available to a wider user community. Data sets that will be made available by 1994 include global-gridded Pliocene and Quaternary sea level; modern average global sea surface temperature and polar sea ice concentration; paleoceanographic data (Sea of Japan); methane release from submarine permafrost (Alaska); glacier velocity measurements (Antarctic); meteorological observations (North Slope of Alaska, Antarctica); fossil pollen data from terrestrial and ocean sediment cores; and eolian surficial deposits on the U.S. Great Plains. Other data sets will be released in future years as data collection and analysis are completed.

*National Biological Survey global change data.* The global change research programs of the Bureau of Land Management, the Fish and Wildlife Service, and the National Park Service have been integrated under the new DOI bureau, the National Biological Survey (NBS). A Global Change Data Center was established in Fort Collins, Colorado, to manage the wide variety of global change data formerly collected by the Bureau of Land Management. Data sets will be made available from studies in meteorology, air quality, soil temperature, surface water chemistry, hydrology, vegetation, squirrel and grasshopper population density, aquatic insects, fire history, tree rings, and pack rat middens. Although some data have been provided to the Global Change Data Center, data sets are also maintained at the project level.

Although NBS global change data management activities are coordinated by the NBS Division of Ecosystem Research, each site is responsible for maintaining, storing, and providing access to its data. Major data sets, formerly collected by National Park Service global change research projects, include multmillennial tree ring growth, Sequoia forest population, and fuel and fire dynamics data in the Sierra Nevada Mountains; tree ring data reflecting water yield along several rivers in the Ozark Mountains dating back to the 12th century; and data from Glacier National Park's Lake McDonald watershed, including physical attributes, vegetation, and AVHRR data. Data sets collected under the Fish and Wildlife Service include tree ring growth records and remote sensing measurements of coastal wetlands of the southeastern states; analysis of sea level changes in the Great Lakes from sediment cores; long-term (25-year) records from the North American Breeding Bird Survey; biology and hydrology of wetlands in the Northern Great Plains; snow cover in the Alaskan tundra from AVHRR data; and fish population, status and trends, and collections.

## GCDIS Access

### Focused Programs

*Global Land Information System.* The USGS has developed the GLIS as an online, computer-based directory and inventory system to provide users with an interactive, user-friendly source for information about the land-related data sets described above. The GLIS is a primary node of the interagency Global Change Data and Information System. Examples of the kinds of data sets accessible through the GLIS are

- Satellite observations (daily AVHRR LAC and GAC, Global Vegetation Index, Normalized Difference Vegetation Index composites from AVHRR LAC data, and Landsat MSS and TM data);
- Land cover/land use data (Mathews Global Vegetation Cover, World Ecosystem Complexes, and Holdridge Life Zones);
- Soils and terrain data (UN-FAO soils map of the world, Zobler soils categories, and SOTER digital data bases); and

- Terrain and cartography data (USGS and NGDC Digital Elevation Models, ETOPO-5 land elevation, USGS Digital Line Graphs, and Digital Chart of the World and elevation data derived from it).

The GLIS users can connect directly to other information systems such as the interagency GCMD, Japan's Earth Observation Information System, the ESA's On-Line Earthnet Data Availability system, and Canada's Global Change Network.

Data set information in the GLIS will be maintained in three levels of detail. The directory level will contain high-level, summary descriptions of entire data sets. Data directory textual searches will be based on discipline, project, sensor key words, geographic location, and other data set parameters. The guide level will contain detailed descriptions of data sets, including information about sensor specifications, extent of coverage, processing history, data quality, and product availability. The inventory level will contain detailed information about individual data sets, such as the time and location of a Landsat scene or AVHRR pass.

*National Biological Survey Science Directory.* The NBS has developed a directory of science data bases that is available in hard-copy form to help users identify and locate terrestrial natural resource data sets. The directory contains summary descriptions of each data set and information about its availability, location, and research staff contact name. Plans include providing this information to the GCMD. The NBS uses an Annual Work Plan to collect, describe, and maintain information about its data sets. This information is available upon request.

### **Contributing Programs**

The mission programs of the DOI bureaus have always included the establishment, maintenance, validation, description, accessibility, and preservation of high-quality, long-term data sets. Many of these data sets are pertinent to global change research; information about these data sets and access to them is provided by the appropriate information systems of the responsible DOI bureaus. In addition, the DOI plays a lead role in interagency coordination activities that relate to many of these data sets. The interagency FGDC was established in 1990 through revised Circular A-16 to coordinate activities related to spatial data, and is chaired by the DOI. Major objectives of revised Circular A-16 are to avoid duplication and minimize costs in mapping and spatial data activities, while maximizing the availability of data to large numbers of users. Also, under the Office of Management and Budget (OMB) Memorandum M-92-01, the DOI is given responsibility for coordinating the water data acquisition and information sharing activities of all Federal agencies, including the quality and quantity of streams, lakes, reservoirs, estuaries, and ground water, plus water use and sedimentation.

*Federal Geographic Data Committee activities.* In the Federal Government, the OMB has identified a need for a "national digital spatial information resource," which has come to be called the National Spatial Data Infrastructure (NSDI). To address this need, the OMB has created the FGDC and has charged it with the responsibility to coordinate various surveying, mapping, and spatial data activities of Federal agencies to meet the needs of the Nation. Executive Order 12906("Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure") was signed by President Clinton on April 11, 1994 to guide the Federal Government's development of the NSDI. The FGDC has also been charged with coordinating a variety of activities with State and local governments and the private sector to develop the NSDI. A series of data-specific subcommittees and crosscutting working groups have been established to address the issues.

NSDI is a framework within which organizations and technology interact to foster more efficient use, management, and production of geospatial data. The NSDI is primarily an umbrella of policies, standards, agreements, and partnerships among a variety of sectors and disciplines that will promote more cost-efficient production, ready availability, and greater use of higher quality geospatial data. The major objectives of the NSDI are to foster enhanced use of geospatial data through better management of existing geospatial data and through more efficient collection and production of new geospatial data in ways that maximize data usefulness for multiple data users. The NSDI will promote development and maintenance of and access to data sets that are needed for national, regional, state, and local analyses. Many of these data sets, developed by agencies to meet ongoing mission responsibilities, are important to global change research. The efforts of the FGDC will improve access to these data sets by global change researchers as well as a wide range of other users.

The FGDC is developing the NSDI to provide a means to document, query, search, and access geospatial data in much the same fashion as is being developed for global change data through the GCDIS. The clearinghouse is based on a distributed approach to geospatial data management that relies on data producers to document their data consistently using the Content Standards for Digital Geospatial Metadata. Data searches will be conducted over electronic networks such as the Internet, using public-domain WAIS query software. Organizations not connected to the network will have access to the clearinghouse through sites such as the Earth Science Information Centers (1-800-USA-MAPS) or through agency-established data servers. A prototype test of the clearinghouse is in progress. A *Manual of Federal Geographic Data Products* was published in hard-copy form (an electronic version is being developed) to provide an initial reference source to the types of spatial data that will become available through the clearinghouse. It contains a comprehensive descriptions of more than 150 publicly distributed Federal geographic data products from 21 Federal agencies, including maps, computer-compatible (digital) data, aerial

photography and multispectral imagery, and other geographically referenced data sets.

*Earth Science Data Directory.* In addition to the GLIS, the USGS maintains the Earth Science Data Directory (ESDD) with more than 2,300 descriptions of data sets from a number of Federal agencies and all 50 States. The ESDD directory-level descriptions provide high-level information about data sources, their size and geographic location, format and media, and technical contacts for each data set.

The USGS also maintains the National Water Data Exchange, a computerized index to more than 450,000 sites for which water data are available from more than 450 organizations.

#### **DOI GCDIS Components Available by April 1, 1995**

*Global Land Information System.* The GLIS will be enhanced to provide full graphical user interface (GUI) capability in the middle of 1994. This capability will include a full suite of X-windows tools for selecting geographical search areas and displaying data coverage and browse files over the Internet. In addition, the XGLIS version will include data set user guides and sample data graphics files that are compatible with the World Wide Web document viewing tools. This will allow XGLIS users to link easily to relevant global change documents hosted by research centers all over the world that are also utilizing World Wide Web.

#### **DOI GCDIS Components Available by April 1, 1996**

*Global Land Information System.* The GLIS will be further expanded to support interactive query and online ordering from the National Digital Cartographic Data Base sales data base (including 1:250,000-scale land use and land cover, 7.5-minute Digital Elevation Models (DEMs), 1:100,000-scale Digital Line Graphs (DLGs), 1-degree DEMs, 1:2,000,000-scale DLGs, and 1:24,000-scale DLGs). This capability will include cross-inventory searching for identifying data sets covering the same geographic area and acquired during the same time period.

The following table represents the data and information access capabilities that will be provided at the EROS Data Center between 1994 and 1996. Other individual DOI GCDIS components will provide a variety of capabilities on various time scales.

**DOI GCDIS Functionality: Access Capability Table**

DOI	1994	1995	1996
Archive	3	2	2
Directory	1	1	1
Guide	1	1	1
Inventory	1	1	1
Browse	2 <sup>1</sup>	2 <sup>1</sup>	2
Order Place	2	2	1
Distribution	2 <sup>1</sup>	2 <sup>1</sup>	1

<sup>1</sup> This level of service available only for selected data sets

**DOI International Linkages Related to Global Change Data**

The UNEP has established its North American GRID facility at the EROS Data Center to provide improved access to global change data and information on an international scale. The EDC was recently designated as the World Data Center for Land Remotely Sensed Data under the auspices of the ICSU. This recognizes its holdings of the world's largest collection of space- and aircraft-acquired imagery of the Earth's land surface, including more than 2 million images acquired from satellites and more than 8 million aerial photographs.

The DOI participates in the CEOS, an international group whose members include NASA, NOAA, the USGS, the ESA, Canada, and Japan. The DOI also participates in the Landsat Ground Station Operators Working Group. Participation by the National Park Service in the U.S. Man and the Biosphere Program facilitates use of the international network of biosphere reserves for comparative ecological studies and sharing of data.

Other DOI bureau global change programs have links to international programs. For example, the USGS is participating in the GEWEX, the GCIP, a part of the WCRP. The USGS is also participating in the Land Cover Change Pilot Study of the IGBP Working Group on Data and Information Systems. Other participating countries include Canada, France, Russia, Australia, and the countries of the European Economic Community.

## **Environmental Protection Agency**

### **Summary**

The EPA has the primary responsibility for data on environmental quality and the distribution and effects of pollutants on human and ecological health. As such, it is both a supplier and consumer of information on the environment, with a large potential for beneficial interchange with the GCDIS that goes well beyond its current role in the USGCRP.

The EPA has relatively small holdings that are derived from the focused USGCRP and more than five terabytes of data that are derived from its regulatory and related programs. For the short term, the EPA will seek agreements with existing GCDIS centers to house its data derived from the focused USGCRP programs. In the longer term, the EPA intends to modernize its potentially contributing data systems in ways compatible with the GCDIS, but the EPA is not currently budgeted to do this at a level that would meet the GCDIS projected schedules. To meet the GCDIS schedules, the EPA will need enhanced funding beyond the current EPA USGCRP budget, as outlined in the USGCRP special issue on data capture (see Appendix A).

If the EPA is successful in obtaining the necessary funding, the EPA intends to develop a full catalog and access system that is interoperable with the GCDIS, maintaining a level of functionality appropriate to the priority of the data assigned by the GCDIS ranking process. Under current funding, the EPA is beginning an effort to develop directory and inventory information for some of its contributing program data, it is developing extensive network and telecommunications connectivity, it is enhancing its supercomputing center activities, and it is developing plans to become a consumer of USGCRP data and information (especially satellite data from the NASA EOSDIS).

The EPA's potentially contributing data result from a broad range of programs aimed at improving human and ecological health, assessing environmental risks, and preventing environmental pollution. Data from the EPA's own programs, as well as data from programs delegated to the states, are stored in a variety of data bases at the EPA's National Computer Center (NCC) in Research Triangle Park, North Carolina. These include

#### *Ambient air quality, source, emissions, and compliance*

- Emissions and compliance from industrial plants,
- Overall status of air quality,
- Reference information (i.e., FIPS codes),

- Emissions from area and mobile sources.

*Ambient water quality, source, emissions, and compliance*

- Public water supplies and compliance with regulations,
- Marine monitoring data pertaining to water quality, oceanographic descriptions, sediment pollutants, physical/chemical characteristics, biological characteristics, and estuary information,
- Permit compliance data,
- Waterway parametric data, ambient, intensive survey, effluent, and biological water quality,
- Specific releases of oils and hazardous substances.

*Hazardous and solid waste*

- Demographics of hazardous waste generation and management,
- Superfund site data,
- Chemical properties, regulatory information, safety data, and toxicity data for emergency response,
- Superfund cleanup records of decision,
- Tracking of generation, transportation, treatment, storage, and disposal of hazardous waste,
- Information on regulated chemicals used in commerce.

*Pesticides and toxic substances*

- Tracking and registration of all pesticide producing establishments,
- Test data used to monitor health, ecological, and safety effects of toxic chemicals used within industries,
- Toxic release inventory.

*Cross-program*

- EPA's civil judicial enforcement activity,
- Environmental monitoring methods index,
- Reference information of facilities,
- Construction grants on waste water treatment,
- Summaries of health risk and regulatory information,

- Online library system,
- Monitoring data for the Nation's ecosystems (Environmental Monitoring and Assessment Program [EMAP]—under development).

**Summary of EPA GCDIS Capabilities Available April 1, 1994**

As of April 1, 1994, the EPA portion of the GCDIS will consist of the components, listed below, accessible at various prototype levels over the Internet through a UNIX public access server at Research Triangle Park, North Carolina. The public access server will be developed as part of the EPA's general strategy to increase access to environmental data and includes efforts focused on the USGCRP as well as other areas.

Extracts from selected program data bases and model results will be made available as part of the Envirofacts data base. The server will support Internet search tools (e.g., Gopher, WAIS, and Mosaic), and contain both data and pointers to data and information. Supported tools will be interoperable with appropriate components of the GCDIS. Electronic versions of *Access EPA*, the *Toxic Release Inventory*, the Guide to Selected National Environmental Statistics in the United States, the Online Library System, and the Information Systems Inventory will also be available on the server to assist users in obtaining EPA information available on line and by hard copy.

Data and information from the focused portion of the USGCRP (e.g., the North American Landscape Characterization data set) will be made available through arrangements with other GCDIS archives. *Contributing* data and information from EPA's regulatory and scientific programs, not derived from the USGCRP activities, will be incorporated into the Envirofacts data base, as funding becomes available and requirements are known from the USGCRP.

**EPA GCDIS Components Available by April 1, 1994**

*The National Computer Center.* The EPA's National Computer Center, located at Research Triangle Park RTP), was founded in 1971 as the Research Triangle Computing Center with a primary focus on EPA's air quality research and enforcement programs. It was renamed the NCC in 1975, to reflect a more nationwide mission. In 1980, all the agency's data processing and telecommunications capabilities were consolidated under the NCC. Today the NCC is a large, standardized, integrated, high-speed computing center with a computing environment that varies from workstations to a large IBM mainframe.

The NCC responds to the demands of new environmental legislation, serves as the central medium for storing and processing information, and provides nationwide

services and support to more than 25,000 Federal, State, and local users of the agency's information resources. Each year it receives information inquiries from these users and solves and documents problems for them. In addition, it supports their hardware, software, telecommunications, and operations needs. The NCC would form the cornerstone of EPA's implementation of the GCDIS and would contain most of the contributing data provided by the agency.

*The National Environmental Supercomputing Center.* The National Environmental Supercomputing Center (NESC) at Bay City, Michigan, is the world's only supercomputer center devoted exclusively to environmental problems. The NESC supports EPA modeling efforts in regional air models, in large ecosystems such as the Chesapeake Bay and Great Lakes, and in computational chemistry for the study of toxic substances.

The NESC provides EPA programs with the ability to simulate environmental systems by mathematical models too large to be executed on conventional computers. The cost and time savings of modeling compared with that for field testing allow EPA to test environmental strategies on the computer without going through the costly and time consuming process of trial and error using the regulatory process. Its role in the GCDIS will be to provide a vehicle for processing environmental models and remote sensing data to create information supporting both research and policy formulation.

*The EPA data communications network.* The EPA's data communications network is designed to provide support to a wide range of clients, including EPA regional offices and laboratories, state and local agencies, and public access. RTP, Washington, D.C., and Cincinnati, Ohio, are connected by a triangle of T1 circuits. Additionally, the NESC in Bay City, Michigan, is connected to RTP with a T3 circuit. The EPA's 10 regional offices are connected by 56-Kbps circuits to both RTP and Cincinnati. All fifty state agencies, as well as Puerto Rico, are connected by analog circuits to the regional offices. Internet connectivity is supported at the T3 level through a connection at Bay City. At the workstation level, the agency supports approximately 19,000 token-ring and 2,500 Ethernet LAN-connected stations. Currently, IBM, VAX, and Cray registration consists of 12,586 users. Virtually all EPA users have the potential for Internet email access through the agency's internal email system, although all employees are not connected to the internal system. The EPA also maintains full Internet connectivity to selected workstations throughout the agency.

*Access EPA.* Published annually since 1991, *Access EPA* is a comprehensive directory and locator that provides descriptive and access information on more than 300 of EPA's major information resources, including data bases, clearinghouses, hotlines, records, models, and documents. Additional information is also available

through the contacts listed. Originally designed to enhance public access to environmental information, this resource will become a key component of the agency's electronic information dissemination effort, especially once it links up with the proposed Government Information Locator Service currently under development.

Printed copies are available in main public and branch libraries and in state environmental libraries. Copies may also be ordered through the Government Printing Office or the National Technical Information Service. A brief quick-reference version, *Access Express*, is available at EPA's Public Information Center. A full-text online version of *Access EPA* is currently accessible through both the EPA Online Library System (OLS) and the Internet.

*Online Library System.* The EPA OLS is a bibliographic data base that references the information resources maintained by the EPA Library Network. It contains bibliographic citations from books, EPA and other Federal agency technical reports, conference proceedings, indexes, audiovisual materials, maps, journals, and a variety of other documents held in the collections of EPA Headquarters and regional and laboratory libraries. The OLS also provides summaries of selected titles. Information contained in the following data bases is also listed in the OLS: the National Catalog, the Hazardous Waste Superfund Collection Database, the Clean Lakes Database, the Office of Pollution Prevention and Toxics Chemical Collection Database, and Regional Files from certain EPA regional libraries.

*Toxic Release Inventory.* The Toxic Release Inventory is a public data base of information on releases of more than 300 toxic chemicals into the environment. The data are self-reported by industry.

*Guide to Selected Environmental Statistics.* The guide is a directory to selected U.S. statistical programs that collect frequently sought after, national-level, time series environmental statistics, compiled and distributed on a regular basis. Each entry describes a separate statistical program, giving program purpose, data coverage, collection methods, geographic coverage, agency contacts, pertinent publications, and data base access options. It is available in both hard-copy and electronic versions, with expanded search features and maps in the electronic version.

Statistical coverage in the guide includes data related to the state of the environment, pressures on the environment, human health and welfare issues, and societal responses to environmental problems. This information was provided by multiple agencies and departments, including EPA.

*Gateway.* Gateway is a Microsoft Windows-based GUI using client-server technology that provides users with a simplified means of accessing and analyzing EPA's integrated data resources. It allows users to access, through pull-down menus

and user-developed reports, agency data resources stored in Envirofacts; to spatially analyze and display that information through a full-function ARC/INFO GIS capability; to employ user-developed analytical tools and models for data analysis; to access worldwide information over the Internet using WAIS and Gopher technology; and to pass the information to desktop tools such as spreadsheets and word processors.

*Envirofacts.* Envirofacts is an integrated, relational data base of subsets of several of EPA's national environmental data bases. It is the relational data base queried by the Gateway user interface. Envirofacts currently contains information from EPA's Facility Index System, Permit Compliance System, Toxic Release Inventory System, and the Superfund information system. In the near future, it is scheduled to include the new Water Quality Storage and Retrieval System water monitoring data base, and work is proceeding on incorporating both the Aerometric Information Reporting System and the Resource Conservation and Recovery Information System data bases. Additional programmatic and monitoring data bases will be included in the future. The Envirofacts data base architecture and structure is currently supported by Version 7.0 of the ORACLE Relational Data Base Management System (RDBMS).

*Information Systems Inventory.* The Information Systems Inventory (ISI) is the definitive source of summary information about approximately 500 of EPA's current information systems, including some models and data bases. The ISI was developed to enhance the agency's ability to track major information systems and to share information across media and program boundaries. Entries under each system in the inventory include system name and acronym, system level, responsible organization, contact person, legislative authority, data base descriptors, access information, hardware and software, system abstract, and key words. The ISI provides searching capabilities and pointer information so that people interested in EPA information can readily determine what EPA systems may be relevant to their interests, and whom to call to get further information.

The ISI is available in hard copy and automated forms in the EPA headquarters and regional libraries, and through the NTIS. Future plans are to incorporate the ISI into an expanded version of Access EPA.

*EMAP Information Management system.* The EMAP Information Management (IM) is the vehicle with which the total EMAP program manages data and information. The EMAP IM will provide the infrastructure for turning raw scientific measurements into useful information products that will serve to guide actions to improve the environment. The EMAP IM is also designed to provide a diverse user community easy access to data and processed information through a series of tools that access data and information over a distributed national network.

The EMAP IM architecture consists of seven conceptual layers: canonical data, virtual repository, security, communication access, tool set, interprocess communications, and user interface. Implementation will progress through proof of concept, technology transfer, and enterprise implementation processes. The goal is to develop by 1997 an environmental information highway allowing analysis of EMAP data across heterogeneous networks of personal and scientific computers and among an interagency federation of users.

*CIESIN catalog.* As part of its efforts to increase public access to environmental data, the EPA has provided assistance to the CIESIN to enable it to develop a prototype catalog of environmental information. The CIESIN will develop directory and inventory portions of an environmental catalog, using selected EPA data bases. These will be maintained on a CIESIN server that is interoperable with the other components of the GCDIS and publicly available over the Internet. Additionally, the CIESIN is developing a prototype Great Lakes Regional Information System, compatible with the Great Lakes Information Network and the GCDIS, that will be accessible by the environmental community over the Internet. The Great Lakes prototype is intended to be the paradigm for other efforts to develop environmental information systems for large geographic regions such as the Chesapeake Bay, the Gulf of Mexico, the U.S.-Mexican Border, and Eastern Europe. The purpose of these efforts, taken together, is to increase the public availability of environmental data and to facilitate human dimensions and Earth science data integration for improved decision making at the local to global level.

#### **EPA GCDIS Components Added by April 1, 1995**

Initially, activity will be focused on developing the Envirofacts data base and constructing the necessary tools and metadata to facilitate access to a prototype information system on the RTP server. It is expected that the CIESIN directory prototype will be sufficiently developed to permit implementation on the RTP server, either as a copy or as a reflector to the CIESIN facility. It is also expected that the Great Lakes and EMAP prototypes will have created access points through the server as well. An online user feedback system will be developed to help improve access and develop content.

#### **EPA GCDIS Components Added by April 1, 1996**

The Envirofacts data base will be expanded to include additional EPA program data with content related to the requirements of the USGCRP. It is expected that the prototypes developed during 1995 will be refined based on the previous year's user feedback. Additional geographic data bases may be added.

**EPA GCDIS Functionality: Access Capability Table\***

EPA	1994	1995	1996
Archive	1	1	1
Directory	3	3	3
Guide	4	3	3
Inventory	4	4	4
Browse	2	2	1
Order Place	2	2	2
Distribution	2	2	1

\* Except for *Archive*, codes refer to the Envirofacts data base and Public Access Server only.

**National Aeronautics and Space Administration****The Ultimate NASA Vision**

NASA's major contributions to the GCDIS are the components of the GCMD and the EOSDIS. The EOSDIS will be a single, distributed system that will serve the Earth science community by archiving and distributing data in support of Earth science research programs. The EOSDIS will provide easy and reliable access to Earth science (e.g., climate, oceanography, land science, hydrology, biogeochemical dynamics) data products from a variety of sources, including the following:

- NASA's series of EOS satellites—a major comprehensive global observing system providing a minimum of 15 years of continuous Earth observations starting with the first launch in 1998.
- Current and future Earth Probe NASA missions, such as the Tropical Rain Measurement Mission (TRMM), the Ocean Topography Experiment (TOPEX/Poseidon), the Upper Atmosphere Research Satellite (UARS), the Total Ozone Mapping Spectrometer (TOMS), the Sea-Viewing Wide-Field of View Sensor (SeaWiFS), and the NASA Scatterometer (NSCAT) on Japan's Advanced EOS.

- International missions in which NASA participates, such as the European Remote-Sensing Satellites (ERS-1 and ERS-2), Japan's Earth Resources Satellite (JERS-1), and Radarsat.
- Field campaigns, such as the Boreal Ecosystem–Atmosphere Study (BOREAS), the Oregon Transect Ecosystem Research (OTTER), and the First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE).
- Existing space-based operations, such as NOAA's AVHRR, TOVS, and GOES, identified as critical to global change research by Pathfinder projects and reprocessed using community consensus algorithms as recommended by designated Science Working Groups.
- Existing data sets, other data essential for the interpretation of measurements, and other activities sponsored by NASA's Mission to Planet Earth program.

In order to protect the significant public investment in data, the EOSDIS has assigned the responsibility for making the data permanently available in an easily usable form to the DAACs and the SEDAC. Along with the SEDAC and the GCMD, the DAACs serve as the primary user interface to the EOSDIS.

In general, there is no period of exclusive access for EOS data. However, some of the data centers hold data sets for which there are specific stipulations that preclude open distribution. Data will be available to scientific users at the marginal cost of filling their requests, and scientific users will be required to make available their research results and any derived data products. Results of significant interest will become part of the EOSDIS data holdings and be distributed under the EOS data policy.

### **Current NASA Capabilities**

The EOSDIS is building on existing discipline-specific Earth science data centers and data systems. The preexisting data systems form the starting point for the EOSDIS. The data systems and data set initiatives that make up today's EOSDIS employ existing electronic networks for data transfer, interactive sessions, and mail. These services are provided through the NASA Science Internet (NSI). This network supports several protocols and is interoperable with the Internet. Currently, the NSI consists of a backbone based on T1 technology (1.5 Mbps) connecting 27 regional networks and more than 100 tail circuits to research sites, reaching approximately 2,500 end users.

### **NASA GCDIS Capabilities Available April 1, 1994**

NASA provides access to Earth science data through several discipline-specific data centers and data systems, most of which can automatically be accessed from the

### EOSDIS DAAC Expertise and Support

Data Center	Concentration	Missions, Instruments
ASF	Sea Ice, Polar Processes	ERS-1, JERS-1, ERS-2, Radarsat
SEDAC	Socioeconomic data	Human dimensions, policymaking applications
NSIDC	Cryosphere and Polar Processes (non-SAR), Cryosphere/Climate Interactions	DORIS, GLAS, MIMR, SSALT, SMMR, SSM/I, TMR
ORNL	Biogeochemical Dynamics, Trace Gas Fluxes,	In situ measurements, Terrestrial/Aquatic/Marine Ecosystem Field Experiments
MSFC	Global Hydrologic Cycle	MIMR, TRMM (TMI, PR, LIS), SSM/I
GSFC	Upper Atmosphere, Atmospheric Dynamics, Global Biosphere, Geophysics	MODIS, AIRS/AMSU/MHS, CZCS, GLAS, HIRDLS, MIMR, MLS, SAGE, SeaWiFS, TOMS TRMM (VIS), UARS
JPL	Physical Oceanography, Air-Sea Interactions	TOPEX/Poseidon, NSCAT, MIMR, Seasat, SSALT, TMR
EDC	Land Processes	ASTER, AVHRR, Landsat, MISR, SAR, airborne sensors
LaRC	Radiation Budget, Clouds, Tropospheric Chemistry	ACRIM, ASTER, EOSP, CERES, ERBE, MOPITT, MISR, EOSP, SAGE, TES, SOLSTICE

GCMD. These centers and systems provide various levels of service for data processing, distribution, and archiving. The numerous disciplines supported include climate, oceanography, land science, hydrology, biogeochemical dynamics, sea ice, geophysics, atmospheric dynamics, radiation budget, and human dimensions.

Researchers may access most of the systems through the Internet and dial-up lines, or visit hard-copy browse facilities. Data are delivered electronically or on standard media, such as 9-track magnetic tape, 8-mm cartridges, or CD-ROMs. System capabilities allow users to search, locate, select, and order products. Searches usually can be limited by geographic area, time, or geophysical parameter. Electronic access is generally free to research users, but a fee may be charged to cover the marginal costs of filling the requests.

### **NASA GCDIS Components Available April 1, 1994**

Although NASA's major contributions to the GCDIS are the GCMD and the EOSDIS, the building blocks available for April 1, 1994, include several other data centers and data systems. These centers and systems are operationally supporting the Earth science research community with various levels of service. Most of these services will be transferred to the EOSDIS data centers as resources allow, although because of funding limitations, a full transition is not envisioned within the next several years. These systems will probably continue to be separately funded to support the research community, and can continue to support the GCDIS at the current access levels.

In addition to the GCMD and the EOSDIS Data Centers described earlier, the following building blocks are available:

- The National Space Science Data Center's (NSSDC) Online Data and Information System Services (NODIS) provides information about past, present, and future space-flight missions and about the investigations flown or scheduled to fly on those missions, detailed information about those data sets primarily archived at NSSDC, and electronic access to a few Earth science data sets, such as TOMS data,
- The Crustal Dynamics Data Information System (CDDIS) maintains information about, stores, and distributes data products acquired to further the advancement of scientific understanding of Earth dynamics, tectonophysics, and earthquake mechanisms, including the results of science support groups analyzing these data sets,
- The Scientific and Technical Information Program's Center for Aerospace Information provides online access to bibliographical data bases and documents.

*Content.* Some of the most frequently requested holdings of each of the data centers and systems contributing to GCDIS on April 1, 1994, are listed in the table to follow. These include several Pathfinder Data Sets. (See *Science Data Plan for the EOS Data and Information System Covering EOSDIS Version 0 and Beyond* for additional details.)

**Major NASA Global Change Holdings on April 1, 1994**

<b>ASF DAAC</b>	Synthetic Aperture Radar (SAR) from the European Space Agency's ERS-1 satellite and the Japanese JERS-1 satellite Level-1 data on digital and photographic media and higher level geophysical products including sea-ice motion, sea-ice classification and ocean wave height and direction (note that restrictions apply based upon prior agreements with international partners) Alaska Landsat imagery dating back to 1972 Alaska AVHRR-HRPT (1-km) data on photographic media since 1974 Alaska High Altitude Aerial Photography (AHAP) on 10-in. film between 1978 and 1986
<b>NSIDC DAAC</b>	Arctic Ocean drifting buoy data Arctic rawinsonde archive Nimbus-7 SMMR Polar Radiances Nimbus-7 SMMR Arctic and Antarctic Sea Ice Concentration DMSP SSM/I Brightness Temperature Grids for the Polar Regions DMSP SSM/I Sea Ice Concentration Grids Navy-NOAA Joint Ice Center Weekly Sea Ice Concentration and Extent—Arctic and Antarctic Greenland (Dye-3) Ice Core Oxygen-18 Versus Depth data Rand Corporation Mean Monthly Global Snow Depth Northern Hemisphere Sea Ice Concentration 1953–1990 DMSP F8 Special Sensor Microwave/Imager (SSM/I) Brightness Temperature Grids
<b>EDC DAAC:</b>	Landsat (MSS, TM, FOLD) NOAA AVHRR LAC and HRPT AVHRR Normal Difference Vegetation Index (NDVI) Air and Space Photography. Airborne Sensor Data Digital Elevation Model Digital Line Graph Data National Uranium Resource Evaluation Data
<b>GSFC DAAC*</b>	Atmosphere and climate data sets from Nimbus, NOAA, GOES geostationary satellites In situ measurements, such as surface station climatologies, and data gathered by coordinated scientific projects. (e.g., the First GARP Global Experiment) TM and MSS images available on 70-mm film rolls, 16-mm film cassettes, or microfiche UARS Level-0 through Level-3 Products CZCS full resolution data (1-km), reduced resolution data (4 km), derived products (4 km), daily global products, 5-day global products, monthly global products (20 km). In situ ship observations from November 1978 to May 1980 covering the mid-Atlantic, the Gulf of Alaska, the Gulf of Mexico, the Northeast coast of Australia, the Dominican Republic, and Spain.

\* Includes contents of the NASA's Climate Data System, Pilot Land Data System, and the CZCS Browse Facility

<b>CDDIS</b>	Analyzed results including precision baseline distances and velocities, Earth rotation and polar motion determinations, length-of-day values, and calculated three-dimensional station positions. Laser ranging data (to both artificial satellites and the Earth's moon) Very Long Baseline Interferometry (VLBI) Global Positioning System (GPS) data starting in 1990 for project-sponsored experiments as well as global networks.
<b>LaRC DAAC</b>	Earth Radiation Budget Experiment Surface Radiation Budget Experiment Stratospheric and Aerosols Gas Experiment International Satellite Cloud Climatology Project First ISCCP Regional Experiment Global Tropospheric Experiment Measurement of Air Pollution from Satellites Stratospheric Aerosol Measurement II RDAT and BANAT
<b>ORNL DAAC</b>	Biogeochemical data from ground-based project that have been funded by NASA (FIFE, OTTER, etc., transferred from the PLDS) Data from continuing and planned field campaigns (e.g., BOREAS), DOE Carbon Dioxide Information and Analysis Center data NSF Long-term Ecological Research data DOE's National Environmental Research Park data U.S. Forest Service forest inventory, forest health monitoring, and watershed research data.
<b>JPL DAAC</b>	AVHRR monthly global MCSST co-registered with CZCS phytoplankton pigment data (MIAMI, GSFC) AVHRR weekly global and regional 18-km MCSST (U. of Miami) TOPEX/Poseidon merged geophysical altimeter data SSM/I global daily ocean wind speed, liquid water, water vapor (Wentz) ATLAST, a PC software atlas and plotting tool for oceanographic sections (Rhines) Ocean Atlas, software atlas and plotting tool for oceanographic sections (Swift, et al.) IMAGIC, an image processing software (Powell, et al.) Tropical Ocean and Global Atmosphere (TOGA) related collection of satellite and in situ data (1985–1990) Geostationary altimeter sea-surface height, SSM/I wind speed, AVHRR SST '87 '89 '90 (Halpern) SSM/I six hourly surface-wind vectors at SSM/I locations and as gridded analysis
<b>MSFC DAAC</b>	SSM/I NESDIS 1b Sensor Counts from F8, F10, F11 SSM/I Wentz Antenna Temperature from F8 and F10 Chang SSM/I Monthly Rain Indices Eischeid Surface Rain Gauge Observations Jaeger Surface Rain Gauge Observations Legates Surface and Shipboard Rain Gauge Observations Morrisey Pacific Atoll Rain Gauge Observations
<b>CIESIN/SEDAC</b>	Public Use Microdata Samples (PUMS) data base HAZDAT data base
<b>NODIS</b>	One year of TOMS data

*Services (access mechanisms).* Most of the data systems can be automatically accessed from the GCMD, which has automated connections to approximately 30 global-change-relevant information systems. Most of the systems, however, require that the user obtain an account before obtaining data. Most of these search and order services can be accessed through the Internet and dial-up connections. Most browse capabilities at this time are limited to hard copy. Some data are available through online services, although most ordered data will be sent on standard media. In 1994, the network backbone will be upgraded to T3-technology (45 Mbps). Contact the user support personnel for the individual systems or data centers as listed in the following table.

#### NASA Earth Science Data Systems Contacts

Component	Electronic Contact	Phone Contact
<b>EOSDIS</b>		
ASF DAAC	rwatabe@santa.asf.alaska.edu	V 907 474-7487 F 907 474-7290
SEDAC	ciesin.info@ciesin.org	V 517 797-2727 F 517 797-2622
NSIDC DAAC	hanson@kryos.colorado.edu	V 303 492-1834 F 303 492-2468
ORNL DAAC	pkk@ornl.gov	V 615 574-7818 F 615 574-4665
MSFC DAAC	msfcuser@microwave.msfc.nasa.gov	V 205 544-4086 F 205 544-5148
GSFC DAAC	daacuso@eosdata.gsfc.nasa.gov	V 301 386-3209 F 301 286-1775
JPL DAAC	podaac@shrimp.jpl.nasa.gov	V 818 354-0906 F 818 393-6720
EDC DAAC	risty@edcserver1.cr.usgs.gov	V 605 594-6507
LaRC DAAC	usersev@eosdis.larc.nasa.gov	V 804 864-8656 F 804 864-8807
<b>Other</b>		
CDDIS	noll@cddis.gsfc.nasa.gov	V 301 286-9283 F 301 286-1776
GCMD	olsen@eosdata.gsfc.nasa.gov	V 301 286-9760
STI CASI	help@sti.nasa.gov	V 301 286-2309
NODIS	request@nssdca.gsfc.nasa.gov	V 301 286-6695

F=Fax V=Voice

*Interoperation.* In this time frame, an integrated implementation of the functions for data searching and ordering across most of the EOSDIS DAACs will be available to a restricted set (about 15) of science advisors.

### **NASA GCDIS Components Available April 1, 1995**

In July 1994, the initial version of EOSDIS will be released—Version 0 (V0). This version is a working prototype with some operational elements, and supports the research of scientific users with currently available data and new data from other missions before the availability of EOS data. It interconnects the existing services of the various EOSDIS DAACs and the SEDAC through electronic networks, interoperable catalogs, and common data distribution procedures to provide better access. With V0, users will have a unified Earth sciences view. Version 0 provides an operating infrastructure to test and refine the evolutionary methodology essential to EOSDIS success.

*Content increases.* More than 250 NASA data sets and approximately 100 related products will be available with the release of EOSDIS V0, including the several Pathfinder products and data products from new missions such as SeaWiFS, Radarsat, and ERS-2. (Note that the distribution of SeaWiFS data will be restricted for 5 years after observation.) See *Science Data Plan for the EOS Data and Information System Covering EOSDIS Version 0 and Beyond* for additional details.

*Service improvements (access mechanisms).* An Information Management System will support data and information search and order across DAACs, SEDAC, and selected non-NASA Affiliated Data Centers. A user will be able to log into any DAAC and find out about the data holdings at other sites. The user will be able to submit a request to order the data from a single site, yet have multiple DAACs fill the order.

Version 0 will employ graphics to provide a simple, effective means of obtaining information about the data sets. Researchers will be able to search for data by geographic area and time period, as well as by such characteristics as data source or geophysical parameters. In addition, users may browse through samples of the data displayed on their computer screens.

*Interoperation.* The most important accomplishment of V0 is that it will link the EOSDIS Data Centers together and provide interoperability among them to give users an Earth science view across the DAACs for searching and ordering data. As a prototype, it will not have all the capabilities, fault tolerance, or reliability of later versions; however, it will support use by the scientific community in day-to-day research activities.

**NASA GCDIS Components Available April 1, 1996**

EOSDIS Version 1 (V1) will be implemented in steps during 1995–1997. It will be designed and developed while V0 is operating, thus benefiting from V0 user feedback.

*Content increases.* ADEOS NSCAT data will be available for distribution from the JPL DAAC. Other data sets will be available from the DAACs based on input from their respective user working groups. In addition, preparations will be made for processing Clouds and Earth's Radiant Energy System and Lightning Imaging Sensor data from TRMM (launched in 1997), archiving and distributing TRMM data, and information management for Landsat-7 (launched in 1998) data.

*Service improvements (access mechanisms).* Version 1 will be physically distributed, but the system will appear completely integrated to users and will provide the complete set of capabilities needed for EOS science and mission operations. Version V1 will be interoperable with version V0.

*Interoperation.* The capabilities of version V1 will progressively replace (after a period of parallel operations) the V0 system to provide a smooth transition to enhanced services.

**NASA GCDIS Functionality: Access Capability Table**

The level of NASA's contribution to GCDIS is indicated in the following table. This table indicates when EOSDIS will have access capabilities analogous to those described for the GCDIS. Although NASA will provide archive functionality through the EOSDIS as noted in the table, NASA has concluded agreements with NOAA and the USGS for long-term archiving of NASA Earth science data.

NASA	1994	1995	1996
Archive	4	4	2–3
Directory	2	2	1
Guide	2–3	2–2	1
Inventory	2	2	1
Browse	2	2	2
Order Place	2	2	1
Distribution	2	2	2

Capabilities currently and for 1994–1995 are being developed within EOSDIS Version 0 (available July 1994). Capabilities in 1996–2000 will be provided by the EOSDIS Versions 1 and 2, built around the EOSDIS Core System (catalog functions available 1996, complete in 1997) and DAAC-unique extensions.

## **National Science Foundation**

### **Summary**

The NSF sponsors a large and diverse research community that both uses and produces global change data and information. Although the NSF has no formal responsibilities for archiving and distributing data and information, it supports a major facility for meteorological, oceanographic, and climatology data sets at NCAR. By and large, however, NSF-supported scientists rely on other Federal agencies for much of their data and information needs and, when appropriate, for archiving and disseminating the research products they produce.

The NSF expects its supported investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections, and other supporting materials created or gathered in the course of the research project. Presently, the enforcement of this requirement varies considerably across the agency. For example, social and behavioral data sets are deposited in an archive for distribution within a year after the completion of a grant. Other discipline divisions at the NSF have no formal policies. The NSF will implement, in collaboration with other GCDIS agencies, a process by which important global change data sets produced with NSF support will be archived, managed, and disseminated for broad community use. Deciding which products are appropriate for this treatment and how the activity will be funded will involve a multilateral process among the research principal investigators, the sponsoring agency program manager, and the appropriate data center and its sponsoring agency. A general rule of thumb for funding responsibility is that the agency program manager should support those activities that are required for the research project itself, and the data center should support those that are required to serve the broader community needs.

### **NSF GCDIS Capabilities Available on April 1, 1994**

*The National Center for Atmospheric Research.* The NCAR Data Support Section (DSS) maintains an archive of more than 400 data sets. Collectively, this archive is designed to support the wide variety of atmospheric and oceanic research at NCAR and within the university community. The complete archive is available to approximately 400 NCAR and 650 university personnel who use the NCAR computing facilities. Access to the data sets is also extended U.S. Government

agencies, the private research sector, and research groups worldwide by distribution on various media and across networks.

The archive is roughly divided into the categories of atmospheric analyses and observations, oceanic analyses and observations, model output, satellite data, and supporting geophysical data sets. The NCAR archive is best known for data sets in the atmospheric and oceanic categories. Groups at NCAR develop some data sets, but a majority are contributed from outside NCAR. Often, the contributed data sets require considerable processing and enhancement to make them useful for large and small computing tasks. Furthermore, the DSS is responsible for preserving the archive, providing access, and redistributing the data sets when needed.

The efficient mass storage system and the connected computing systems provide the DSS with the capability to routinely manipulate large amounts of data. This data facility contributes important data support for research projects at NCAR and throughout the research community. For example, NCAR prepares data for Global Atmospheric Reanalysis, the Comprehensive Ocean-Atmosphere Data Set is produced in cooperation with NOAA, and output from climate models is collected and distributed worldwide for assessment studies. Continued work on these and many other large data manipulation efforts will be a primary focus for the DSS in years to come.

Access to information about data sets and data itself is available through published documents, the DSS support staff, and information systems connected to the Internet. Since 1992, persons with Internet connectivity have been able to access and browse an online data information system at NCAR. This system (using ftp and a simple directory structure), while very successful, has been enhanced to also support World Wide Web information searches through the National Center for Supercomputing Applications Mosaic interface. Mosaic supports several common GUIs for browsing hypertext and simple text files. During 1993 users connected to the system about 2,800 times each month, obtaining about 7,000 files each month. From the files copied, about 2,200 are directly related to specific information and documentation about data sets within the archive. The combination of the simple ftp and GUI systems for viewing information about NCAR's data holdings will continue to be developed in the future.

Not only does NCAR provide data support and distribution on a per request basis, but it has also begun a project to produce CD-ROMs for mass distribution. The data to be placed on the CD-ROMs will first be selected from the atmospheric analyses and observations categories. One CD-ROM was completed last year, several are in the planning phase for completion this year, and others will follow in later years. Presently, NCAR has several data sets that need to be distributed, but are too large

for CD-ROMs or the existing network communications. NCAR intends to distribute these data on existing and new low-cost media as they are developed.

The NCAR archives are diverse in type and sometimes large in volume; so far, the data support focus has been on data content, data information, and easy user data access. NCAR will remain committed to these principles.

Access to data and information about data at NCAR will increase as the aforementioned projects progress. More data will be available on CD-ROMs, and growth of the Internet-accessible information system will continue to take advantage of the most practical and publicly available information distribution tools. Many of the NCAR data sets are presently described in the GCMD. The NCAR will continue to provide information to the directory, as well as support automatic connection to the NCAR data information system. Basic descriptions of all data sets are electronically available now. This information resource will be strengthened as the descriptions are enhanced and more inventories and browse data are made available. This system will meet the level four contribution criteria for GCDIS directory, guide, and inventory functionality.

*The Inter-University Consortium for Political and Social Research.* The NSF does not directly administer an archive of data relevant to the human dimensions of global change. However, the NSF has contributed to the support of the primary such archive in the United States—the Inter-University Consortium for Political and Social Research (ICPSR), based at the University of Michigan. The ICPSR was established in 1962, and is funded primarily from fees paid by more than 370 member academic institutions in more than two dozen countries. Most social and behavioral science data sets collected on the basis of NSF grants are deposited in the ICPSR within a year after the completion of the grant.

The bulk of data sets archived in the ICPSR consists of survey data on individual attitudes and behavior collected by scholars. The archive also contains U.S. census data and surveys on population, health, labor, income, and other conditions collected by Government agencies, international organizations, and other research institutions. The ICPSR has been working with the CIESIN to identify those data sets in the archive that would be of interest to global change researchers in order to make those available through the GCDIS. Negotiations with the CIESIN, with providers of data, and with the ICPSR member institutions have to be completed before it will be clear what the extent and exact form of ICPSR participation in the GCDIS will take. However, it is anticipated that by mid-1994 a substantial portion of the data archived at the ICPSR relevant to global change research will be available through the GCDIS.

Modes of access to the ICPSR have been in transition for the last several years. Traditionally, users could scan a list of holdings and order data sets on magnetic tape

accompanied by hard-copy codebooks. In recent years, the ICPSR has made an increasing number of data sets available in alternative formats—on disk, CD-ROM, or by ftp transmission. Initially, the available format of the data the ICPSR prepares for access through the GCDIS will depend on the current state of the particular data set in the general ICPSR archive. The ICPSR is working with both the CIESIN and the National Center for Supercomputer Applications (NCSA) to develop a high level of online access to ICPSR data sets through the MOSAIC interface being developed by NCSA. By the end of 1994, the ICPSR plans to have one large survey available as an experiment within this environment. Completion of the transition to this level of access for other data sets in the ICPSR archive will require a number of years and substantial funding.

#### **NSF GCDIS Data Available by 1995–96**

*National Center for Atmospheric Research.* As the transmission capacity of networks increases, larger amounts of data will be directly transferred to and from NCAR. Accompanying the growth of network capability will be the development of network tools that efficiently search for data, browse data, and examine and display information files. The resident NCAR data information system will remain flexible and capable of adapting to the new network features and, in doing so, will meet the higher levels of interoperability and connectivity required for the GCDIS.

Since NCAR has long been a data provider to the university community and worldwide, it presently offers level-3 and -2 GCDIS data order and distribution functionality, respectively. As more specific GCDIS guidelines are defined, NCAR will make procedural adjustments to accommodate these standards as considered appropriate.

#### **NCAR GCDIS Functionality: Access Capability Table**

NCAR	1994	1995	1996
Archive	2-4	2-4	2-4
Directory	4	4	4
Guide	2-4	2-4	2-4
Inventory	4	4	4
Browse	3-4	3-4	3-4
Order Place	3	3	3
Distribution	2	1-2	1-2

*The Inter-University Consortium for Political and Social Research.* For 1995–96, work can be expected to continue both on making an increasing number of the relevant ICPSR data sets available through the GCDIS and on upgrading the access levels to those data sets that are already available.

*The Long-Term Ecological Research Program.* Centralized and local (i.e., project level) data management activities of NSF's LTER currently meet or exceed the GCDIS minimum requirements. Although the LTER is not formally a USGCRP-supported activity, it does serve as a good example of the direction and intent of the NSF with regard to broad accessibility of major ecological data sets. At this time, openly accessible LTER data sets include mostly physical environmental data, but there are plans for broader incorporation of primary biological and ecological data as well.

## **U.S. Department of Agriculture**

### **Summary**

The USDA plays an important dual role in global change research and other programs related to the environment and natural resources. USDA agencies support agriculture, forestry, and ranching programs that ensure a continued and healthy food supply for the Nation and our trading partners throughout the world. At the same time, the USDA has a vital interest in and has responsibilities for the protection of the land to preserve it for future generations.

In carrying out its various missions, the USDA conducts research to better sustain agriculture and forestry in the United States and to provide technical assistance in managing private and public lands. Agriculture, forest, and grazing management are affected by weather, climate, and other environmental variables. USDA's global change research program focuses on understanding how global change factors affect U.S. and international forest, pastoral, and agricultural ecosystems, and how these ecosystems affect the environmental variables that contribute to global change.

USDA agencies maintain various centers of data and information that have been developed to support the mission programs of each agency. These programs include the areas of ecological systems and dynamics, biochemical dynamics, climate and hydrological systems, and human interactions with land use. The USDA is in a unique position of having substantial long-term data and information relevant to global change and terrestrial ecosystems. The vast majority of USDA data and information of significant value to global change research were collected to support other agricultural programs, however, and do not carry an easily identified global change label. An intensive effort will be required to capture and incorporate these data and information into the global change research program.

At present, the USDA has no plans to develop a centralized data center to maintain or archive global change data and information. The emphasis will be on providing standards for collecting and managing data—while leaving data with the collecting agencies—and on coordination to promote effective access to these valuable information resources. The goal is to maintain distributed data that will appear fully integrated to the user.

#### **USDA GCDIS Components Available by April 1, 1994**

The incorporation of USDA data and information holdings into the GCDIS will evolve during the coming years. Resources must be allocated to identify and describe priority data holdings and perform any modifications to these holdings to enhance the total knowledge base upon which policymakers and others must make critical decisions regarding land use, agriculture, and forestry issues.

The National Agricultural Library (NAL) has played a major role in identifying, cataloging, and providing access to worldwide published information related to global change issues. The NAL provides a continuing and expanding focal point for identifying and finding desired data and information.

The Current Research Information System is available through the Internet and commercial information distribution systems. This system provides information on the status of research projects funded by the Cooperative State Research Service of the USDA. The USDA has initiated several new projects to support the need to make global-change-related data and information readily available to its various user communities. A key first step is the collection of metadata describing the data sets and other information products held at various locations throughout the country. This inventory includes special information required to define spatial data and models and the particular data sets required for operation of the models. Data bases inventoried are being categorized as weather data (e.g., temperature, precipitation), atmospheric data (trace gases, deposition), soils data, forest data, plant and vegetation data, animal data, pest data, hydrologic data, economic data, and more.

A second follow-on project is the implementation of an automated locator and directory system to house the inventory metadata. This system will be interoperable with the GCMD and other servers, such as the WAIS servers installed at other Federal locations. This locator is not expected to be operational before April 1994.

The USDA has installed a department-wide Internet node for access by all of its agencies to Internet resources. This will supplement or replace several agency Internet connections now in place and will provide outside access to the USDA global change locator and directory.

A project is underway to identify and bring attention to important data sets and information that are at risk of being lost, either because of the poor quality of the storage media or the loss of knowledge of the data due to lack of appropriate documentation or the unavailability of the principal investigator who understood the data. This is a long-term effort, which is expected to result in the development of new policies and a changed culture to prevent this type of loss from seriously degrading future archiving efforts.

Snow, precipitation, air temperature, other climate parameters, and soil moisture and temperature are collected by the Soil Conservation Service (SCS) and automated under a global change pilot project. This Snow Telemetry System is enhanced through manual surveys to provide water supply forecasting capability. These data sets are easily accessible to the public through the SCS Centralized Forecasting System maintained at the Climate Data Access Facility at Portland, Oregon. This historical data set, largely of the western United States, is updated annually.

The USDA maintains several other important centers of data that currently are not readily accessible to the public. These include several maintained by the SCS. The National Cartography and Geographic Information Systems Center in Fort Worth, Texas, contains cartographic data bases, maps, selected digitized hydrologic units, soils geographic data, and orthophotography. The Iowa State University maintains the soils tabular data as well as the National Resources Inventory, which includes soil characteristics, land cover and use, erosion, land treatment needs, vegetative condition, and potential for cropland conversion. The plants data base is maintained at Fort Collins, Colorado, and various other national data sets are maintained at Washington, D.C.

Access to these data sets by researchers or others is only by request. Significant efforts are planned to improve access, but major resources to make this possible will be required.

Aerial photography maps are available through the USDA ASCS Aerial Photography Field Office in Salt Lake City. These photographs are used for conservation practices, locating field boundaries, tax assessment, urban development, pollution studies, and watershed studies. Maps of all National Forests and National Grasslands are available on paper from the USDA Forest Service in Salt Lake City.

The USDA is coordinating an interagency effort to define and develop a thesaurus for global change research data and information. The pilot system uses text imaging and retrieval software that provides word meanings and relationships derived from dictionaries and thesauruses. The pilot system will allow easy access and record accuracy to global change data and information through the Internet using the Z39.50 standard for information search and retrieval.

**USDA GCDIS Components Added by April 1, 1995.**

The inventory and locator effort will continue as more data sets become available. The locator will be accessible through networks to any member of the public. The contents will be enhanced and the capabilities for better understanding of the specific data sets will be improved. Knowledge of the data is only the beginning step, however. Many of these data sets will need to be relocated or formatted to allow easy online access. This will be an evolving process, and progress will be based on priorities and available resources.

Data standards relating to such aspects as collection techniques, automation approach, and element definitions and naming conventions represent a critical parallel effort to those described. Several projects are in their initial stages and are progressing more slowly than desired because of limited resources. Some are being done in conjunction with other organizations such as the FGDC, and many involve State and local government cooperative projects.

Specific data holdings to be made available and access mechanisms to be used cannot be defined at this time.



## C. List of Acronyms and Abbreviations

3DNEPH	Three-Dimensional Nephanalysis
AAAS	American Association for the Advancement of Science
ADC	Affiliated Data Center
ADEOS	Advanced Earth Observing System
AFGWC	Air Force Global Weather Central
ARCSS	Arctic Systems Sciences Program
ARM	Atmospheric Radiation Measurement Program
ASCII	American Standards Committee for Information Interchange
ASF	Alaska SAR Facility
AVHRR	Advanced Very High Resolution Radiometer
BOREAS	Boreal Ecosystem–Atmosphere Study
CARDS	Comprehensive Aerological Research Data Set
CART	Cloud and Radiative Testbed
CD-ROM	Compact Disk–Read Only Memory
CDDIS	Crustal Dynamics Data Information System
CDIAC	Carbon Dioxide Information Analysis Center
CEES	Committee on Earth and Environmental Sciences
CENR	Committee on Environment and Natural Resources Research
CEOS	Committee on Earth Observations Satellites
CEOS-IDN	CEOS International Directory Network
CEOS-WGD	CEOS Working Group on Data
CGED	Committee on Geophysical and Environmental Data
CIESIN	Consortium for International Earth Sciences Information Network
CIR	Center for International Research
CIRLA	Cloud Information Reference Archive and Library
CMDL	Climate Modeling and Diagnostics Library
CO <sub>2</sub>	Carbon Dioxide
COADS	Comprehensive Ocean Atmosphere Data Set
CODATA	Committee on Data for Science and Technology (ICSU)
CSA	Canadian Space Agency
CZCS	Coastal Zone Color Scanner
DAAC	Distributed Active Archive Center
DCP	Data Collection Platform
DEM	Digital Elevation Model
DIF	Directory Interchange Format
DLG	Digital Line Graph
DMA	Defense Mapping Agency
DMSP	Defense Meteorological Satellite Program
DOC	Department of Commerce
DOD	Department of Defense
DOE	Department of Energy
DOI	Department of the Interior
DSS	Data Support Section (NSF)

DTIC	Defense Technical Information Center
EDC	EROS Data Center
EIA	Energy Information Administration
EMAP	Environmental Monitoring and Assessment Program
EO-ICWG	Earth Observation International Coordination Working Group
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
EPA	Environmental Protection Agency
EROS	Earth Resources Observation Systems
ERS	European Remote-Sensing Satellite
ESA	European Space Agency
ESDD	Earth Science Data Directory
ESnet	Energy Sciences Network
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FAO	Food and Agriculture Organization (United Nations)
FDDI	Fiber-Optic Digital Data Interface
FGDC	Federal Geographic Data Committee
FIFE	First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment
FIPS	Federal Information Processing Standards
FNMOC	Fleet Numerical Meteorology and Oceanography Center
ftp	File Transfer Protocol
FY	Fiscal Year
GAC	Global Area Coverage
GCCS	Global Change Catalog System
GCDIS	Global Change Data and Information System
GCH	Global Change Climate and History
GCIP	Global Continental International Project
GCM	General Circulation Model
GCMD	Global Change Master Directory
GCOS	Global Climate Observing System
GCRCIO	Global Change Research Information Office
GEWEX	Global Energy and Water Cycle Experiment
GHCN	Global Historical Climate Network
GHG	Greenhouse Gas
GILS	Government Information Locator Service
GIS	Geographic Information System
GLIS	Global Land Information System
GLOED	Global Emissions Database
GNMDF	Global Network MetaData File
GOES	Geostationary Operational Environmental Satellite
GOOS	Global Ocean Observing System
GRID	Global Resource Information Database (UNEP)
GSFC	Goddard Space Flight Center
GTCP	Global Tropospheric Chemistry Program
GUI	Graphical User Interface
HDP	Human Dimensions of Global Environmental Change Programme
HDP-DIS	Human Dimensions of Global Environmental Change Programme Data and Information System

HIRAS	High Resolution Analysis System
HPCC	High Performance Computing and Communications
IAC	Information Analysis Center
ICPSR	Inter-University Consortium for Political and Social Research
ICSU	International Council of Scientific Unions
IDN	International Directory Network
IEOS	International Earth Observing System
IGBP	International Geosphere–Biosphere Programme
IGBP-DIS	International Geosphere–Biosphere Programme Data and Information System
IM	Information Management
IOC	Intergovernmental Oceanographic Commission
IODE	International Oceanographic Data and Information Exchange
IPCC	Intergovernmental Panel on Climate Change
IRIS	Division of Information, Robotics, and Intelligent Systems (National Science Foundation)
ISCCP	International Satellite Cloud Climate Project
ISI	Information Systems Inventory
ISLSCP	International Satellite Land Surface Climatology Project
ISO	International Standards Organization
ISO/OSI	International Standards Organization/Open Systems Interconnections
ISSC	International Social Science Council
IWGDMGC	Interagency Working Group on Data Management for Global Change
JERS-1	Japan's Earth Resources Satellite
JGOFS	Joint Global Ocean Flux Study
JIC	Joint Ice Center
JPL	Jet Propulsion Laboratory
Kbps	Kilobits per second
LAN	Local Area Network
LaRC	Langley Research Center
LTER	Long-Term Ecological Research
LTER	Long-Term Ecological Research
MAN	Metropolitan Area Network
Mbps	Megabits per second
MC&G	Mapping, Charting, and Geodesy
MD	Master Directory
METAR	Meteorological Aviation Report
MITI	Ministry of International Trade and Industry
MODIS	Moderate Resolution Imaging Spectrometer
MSFC	Marshall Space Flight Center
NAL	National Agricultural Library
NALC	North American Landscape Characterization
NARA	National Archive and Records Administration
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NASDA	National Space Development Agency (Japan)
NBS	National Biological Survey
NCAR	National Center for Atmospheric Research
NCC	National Computer Center
NCDC	National Climatic Data Center

NCSA	National Center for Supercomputer Applications
NCSEB	National Center for Synthesis in Experimental Biology
NDC	National Data Center
NEDRES	National Environmental Data Referral Service
NEIC	National Energy Information Center
NESC	National Environmental Supercomputing Center
NESDIS	National Environmental Satellite, Data, and Information Service
NEW	NOAA Environmental Watch
NEXRAD	Next Generation Weather Radar
NGDC	National Geophysical Data Center
NIC	Network Information Center
NII	National Information Infrastructure
NIST	National Institute of Standards and Technology
NMFS	National Marine Fisheries Service
NOAA	National Oceanic and Atmospheric Administration
NOAADIR	NOAA Environmental Services Data Directory
NODC	National Oceanographic Data Center
NODIS	NSSDC Online Data and Information System Services
NOS	National Ocean Service
NREN	National Research and Education Network
NSCAT	NASA Scatterometer
NSDI	National Spatial Data Infrastructure
NSF	National Science Foundation
NSFNET	National Science Foundation Network
NSI	NASA Science Internet
NSIDC	National Snow and Ice Data Center
NSN	NASA Science Network
NSSDC	The National Space Science Data Center
NTIS	National Technical Information Service
OES	Bureau of Oceans and International Environmental and Scientific Affairs
OL-A	Operating Location-A
OLS	Online Library System
OMB	Office of Management and Budget
ONR	Office of Naval Research
ORNL	Oak Ridge National Laboratory
OSTI	Office of Scientific and Technical Information
OTA	Office of Technology Assessment
OTTER	Oregon Transect Ecosystem Research
PCMDI	Program for Climate Model Diagnosis and Intercomparison
PIBAL	Pilot Balloon
PIC	Product Information Catalog
POEM	Polar-Orbit Earth Observation Mission
POSIX	Portable Operating System Interface
RDBMS	Relational Data Base Management System
RIC	Regional Information Centre (IGBP)
RTNEPH	Real Time Nephanalysis
SAA	Satellite Active Archive
SAMS	Surface Automated Meteorological Systems
SAR	Synthetic Aperture Radar

SCS	Soil Conservation Service
SDSD	Satellite Data Services Division
SDTS	Spatial Data Transfer Standard
SeaWiFS	Sea-Viewing Wide-Field of View Sensor
SEDAC	Socioeconomic Data Center
SGCR	Subcommittee on Global Change Research
SGCR	Subcommittee on Global Change Research
SPN	Shared Processing Network
SSM/I	Special Sensor Microwave/Imager
STA	Science and Technology Agency
STD	Salinity/Temperature/Depth
STI	Scientific and Technical Information
TCP/IP	Transmission Control Protocol/Internet Protocol
TOGA	Tropical Ocean Global Atmosphere Experiment
TOMS	Total Ozone Mapping Spectrometer
TOPEX/Poseidon	Ocean Topography Experiment
TOVS	Tiros Operational Vertical Sounder
TRMM	Tropical Rainfall Measuring Mission
UARS	Upper Atmosphere Research Satellite
UCAR	University Corporation for Atmospheric Research
UNEP	United Nations Environment Programme
UNIDATA	University Data System
USAF	U.S. Air Force
USDA	U.S. Department of Agriculture
USGCRP	U.S. Global Change Research Program
USGS	U.S. Geological Survey
USMARC	U.S. Machine Readable Cataloging
WAIS	Wide-Area Information Server
WAN	Wide Area Network
WCRP	World Climate Research Programme
WDC	World Data Center
WGD	Working Group on Data
WMO	World Meteorological Organization
WOCE	World Ocean Circulation Experiment